

Express Mail Label No. ET176332310US
Date of Deposit: February 25, 2002

PATENT APPLICATION
Attorney Docket No. 21402-279 (Cura-579)

PROTEINS AND NUCLEIC ACIDS ENCODING SAME

RELATED APPLICATIONS

JC828 U.S. PTO
10/085198
02/25/02

This application claims priority from U.S.S.N. 60/271,646, filed February 26, 2001 (21402-279); U.S.S.N. 60/276,401, filed March 16, 2001 (21402-279 D1); U.S.S.N. 60/311,981, filed August 13, 2001 (21402-279 IFC-01); U.S.S.N. 60/312,858, filed August 16, 2001 (21402-279 IFC-02); U.S.S.N. 60/271,840, filed February 27, 2001 (21402-280); U.S.S.N. 60/277,324, filed March 20, 2001 (21402-280 D1); U.S.S.N. 60/286,096, filed April 24, 2001 (21402-280 A1); U.S.S.N. 60/299,695, filed June 20, 2001 (21402-280 IFC-01); U.S.S.N. 60/315,614, filed August 29, 2001 (21402-280 IFC-02); U.S.S.N. 60/272,405, filed February 28, 2001 (21402-281); U.S.S.N. 60/272,410, filed February 28, 2001 (21402-282); U.S.S.N. 60/272,414, filed February 28, 2001 (21402-283); U.S.S.N. 60/278,660, filed March 20, 2001 (21402-283A); U.S.S.N. 60/280,234, filed March 30, 2001 (21402-283B); U.S.S.N. 60/272,404, filed February 28, 2001 (21402-284); U.S.S.N. 60/280,039, filed March 30, 2001 (21402-284 B1); U.S.S.N. 60/313,280, filed August 17, 2001 (21402-284 IFC-01); U.S.S.N. 60/322,818, filed September 17, 2001 (21402-284 C1); U.S.S.N. 60/273,300, filed March 2, 2001 (21402-286); U.S.S.N. 60/280,818, filed April 2, 2001 (21402-286 H1); U.S.S.N. 60/288,353, filed May 3, 2001 (21402-286 J1); U.S.S.N. 60/294,834, filed May 31, 2001 (21402-286 H2); U.S.S.N. 60/299,845, filed June 21, 2001 (21402-286 IFC-01); U.S.S.N. 60/272,922, filed March 2, 2001 (21402-287); U.S.S.N. 60/272,787, filed March 2, 2001 (21402-288); U.S.S.N. 60/285,754, filed April 23, 2001 (21402-288 B1); U.S.S.N. 60/303,242, filed July 5, 2001 (21402-288 D1); U.S.S.N. 60/273,048, filed March 2, 2001 (21402-289); U.S.S.N. 60/283,443, filed April 12, 2001 (21402-289 A); and U.S.S.N. 60/291,703, filed May 17, 2001 (21402-289 B); each of which is incorporated by reference in their entirety.

BACKGROUND OF THE INVENTION

The invention generally relates to nucleic acids and polypeptides encoded therefrom. More specifically, the invention relates to nucleic acids encoding cytoplasmic, nuclear, membrane bound, and secreted polypeptides, as well as vectors, host cells, antibodies, and recombinant methods for producing these nucleic acids and polypeptides.

SUMMARY OF THE INVENTION

The invention is based in part upon the discovery of nucleic acid sequences encoding novel polypeptides. The novel nucleic acids and polypeptides are referred to herein as NOVX, or NOV1-NOV91 nucleic acids and polypeptides. These nucleic acids and polypeptides, as well as derivatives, homologs, analogs and fragments thereof, will hereinafter be collectively designated as “NOVX” nucleic acid or polypeptide sequences.

In one aspect, the invention provides an isolated NOVX nucleic acid molecule encoding a NOVX polypeptide that includes a nucleic acid sequence that has identity to the nucleic acids disclosed in SEQ ID NOS: 2*n*-1, wherein *n* is any integer between 1 and 107. In some embodiments, the NOVX nucleic acid molecule will hybridize under stringent conditions to a nucleic acid sequence complementary to a nucleic acid molecule that includes a protein-coding sequence of a NOVX nucleic acid sequence. The invention also includes an isolated nucleic acid that encodes a NOVX polypeptide, or a fragment, homolog, analog or derivative thereof. For example, the nucleic acid can encode a polypeptide at least 80% identical to a polypeptide comprising the amino acid sequences of SEQ ID NOS: 2*n*, where *n* is any integer between 1 and 107. The nucleic acid can be, for example, a genomic DNA fragment or a cDNA molecule that includes the nucleic acid sequence of any of SEQ ID NOS: 2*n*-1.

Also included in the invention is an oligonucleotide, *e.g.*, an oligonucleotide which includes at least 6 contiguous nucleotides of a NOVX nucleic acid (*e.g.*, SEQ ID NOS:2*n*-1) or a complement of said oligonucleotide.

Also included in the invention are substantially purified NOVX polypeptides (SEQ ID NOS:2n). In certain embodiments, the NOVX polypeptides include an amino acid sequence that is substantially identical to the amino acid sequence of a human NOVX polypeptide.

The invention also features antibodies that immunoselectively bind to NOVX polypeptides, or fragments, homologs, analogs or derivatives thereof.

In another aspect, the invention includes pharmaceutical compositions that include therapeutically- or prophylactically-effective amounts of a therapeutic and a pharmaceutically-acceptable carrier. The therapeutic can be, *e.g.*, a NOVX nucleic acid, a NOVX polypeptide, or an antibody specific for a NOVX polypeptide. In a further aspect, the invention includes, in one or more containers, a therapeutically- or prophylactically-effective amount of this pharmaceutical composition.

In a further aspect, the invention includes a method of producing a polypeptide by culturing a cell that includes a NOVX nucleic acid, under conditions allowing for expression of the NOVX polypeptide encoded by the DNA. If desired, the NOVX polypeptide can then be recovered.

5 In another aspect, the invention includes a method of detecting the presence of a NOVX polypeptide in a sample. In the method, a sample is contacted with a compound that selectively binds to the polypeptide under conditions allowing for formation of a complex between the polypeptide and the compound. The complex is detected, if present, thereby identifying the NOVX polypeptide within the sample.

10 The invention also includes methods to identify specific cell or tissue types based on their expression of a NOVX.

Also included in the invention is a method of detecting the presence of a NOVX nucleic acid molecule in a sample by contacting the sample with a NOVX nucleic acid probe or primer, and detecting whether the nucleic acid probe or primer bound to a NOVX nucleic acid molecule in the sample.

15 In a further aspect, the invention provides a method for modulating the activity of a NOVX polypeptide by contacting a cell sample that includes the NOVX polypeptide with a compound that binds to the NOVX polypeptide in an amount sufficient to modulate the activity of said polypeptide. The compound can be, *e.g.*, a small molecule, such as a nucleic acid, peptide, polypeptide, peptidomimetic, carbohydrate, lipid or other organic (carbon containing) or inorganic molecule, as further described herein.

Also within the scope of the invention is the use of a therapeutic in the manufacture of a medicament for treating or preventing disorders or syndromes including, *e.g.*, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial
25 septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, hypercoagulation, hemophilia, idiopathic thrombocytopenic purpura, heart failure, secondary pathologies caused by heart failure and hypertension, hypotension, angina pectoris, myocardial infarction, tuberous sclerosis, scleroderma, transplantation, autoimmune disease, lupus erythematosus,
30 viral/bacterial/parasitic infections, multiple sclerosis, autoimmune disease, allergies, immunodeficiencies, graft versus host disease, asthma, emphysema, ARDS, inflammation and modulation of the immune response, viral pathogenesis, aging-related disorders, Th1 inflammatory diseases such as rheumatoid arthritis, multiple sclerosis, inflammatory bowel diseases, AIDS, wound repair, obesity, diabetes, endocrine disorders, anorexia, bulimia, renal

artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic, renal tubular acidosis, IgA nephropathy, nephrological diseases, hypercalcaemia, Lesch-Nyhan syndrome, Von Hippel-Lindau (VHL) syndrome, trauma, regeneration (in vitro and in vivo), Hirschsprung's disease, Crohn's Disease, appendicitis, endometriosis, laryngitis, psoriasis, actinic keratosis, acne, hair growth/loss, alopecia, pigmentation disorders, myasthenia gravis, alpha-mannosidosis, beta-mannosidosis, other storage disorders, peroxisomal disorders such as Zellweger syndrome, infantile Refsum disease, rhizomelic chondrodysplasia (chondrodysplasia punctata, rhizomelic), and hyperpyruvic acidemia, osteoporosis, muscle disorders, urinary retention, Albright Hereditary Osteodystrophy, ulcers, Alzheimer's disease, stroke, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, behavioral disorders, addiction, anxiety, pain, neuroprotection, Stroke, Aphakia, neurodegenerative disorders, neurologic disorders, developmental defects, conditions associated with the role of GRK2 in brain and in the regulation of chemokine receptors, encephalomyelitis, anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, Gilles de la Tourette syndrome, leukodystrophies, cancers, breast cancer, CNS cancer, colon cancer, gastric cancer, lung cancer, melanoma, ovarian cancer, pancreatic cancer, kidney cancer, colon cancer, prostate cancer, neuroblastoma, and cervical cancer, Neoplasm; adenocarcinoma, lymphoma; uterus cancer, benign prostatic hypertrophy, fertility, control of growth and development/differentiation related functions such as but not limited maturation, lactation and puberty, reproductive malfunction, and/or other pathologies and disorders of the like.

The therapeutic can be, *e.g.*, a NOVX nucleic acid, a NOVX polypeptide, or a NOVX-specific antibody, or biologically-active derivatives or fragments thereof.

For example, the compositions of the present invention will have efficacy for treatment of patients suffering from the diseases and disorders disclosed above and/or other pathologies and disorders of the like. The polypeptides can be used as immunogens to produce antibodies specific for the invention, and as vaccines. They can also be used to screen for potential agonist and antagonist compounds. For example, a cDNA encoding NOVX may be useful in gene therapy, and NOVX may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from the diseases and disorders disclosed above and/or other pathologies and disorders of the like.

5

10

20

30

preferred embodiments, the disorder, includes, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like.

In yet another aspect, the invention can be used in a method to identify the cellular receptors and downstream effectors of the invention by any one of a number of techniques commonly employed in the art. These include but are not limited to the two-hybrid system, 5 affinity purification, co-precipitation with antibodies or other specific-interacting molecules.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

15 Other features and advantages of the invention will be apparent from the following detailed description and claims.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides novel nucleotides and polypeptides encoded thereby.

20 Included in the invention are the novel nucleic acid sequences and their encoded polypeptides. The sequences are collectively referred to herein as "NOVX nucleic acids" or "NOVX polynucleotides" and the corresponding encoded polypeptides are referred to as "NOVX polypeptides" or "NOVX proteins." Unless indicated otherwise, "NOVX" is meant to refer to any of the novel sequences disclosed herein. Table A provides a summary of the NOVX

25 nucleic acids and their encoded polypeptides.

TABLE A. Sequences and Corresponding SEQ ID Numbers

NOVX Assignment	Internal Identification	SEQ ID NO (nucleic acid)	SEQ ID NO (polypeptide)	Homology
1	CG57602-01	1	2	DJ0751H13.1 Protein
2	CG57558-01	3	4	Mac25/IGFBP7
3	CG57560-01	5	6	Calmodulin Binding Protein Kinase
4A	CG57547-01	7	8	TRANSIENT RECEPTOR POTENTIAL-RELATED PROTEIN
	CG57547-02			TRANSIENT RECEPTOR

45A	CG57656-01	105	106	Ig/Fibronectin domain
45B	CG57656-02	107	108	Ig/Fibronectin domain
46	CG57682-01	109	110	G2/MITOTIC-SPECIFIC CYCLIN B2
47	CG57764-01	111	112	ALR
48	CG57713-01	113	114	SODIUM/BILE ACID COTRANSPORTER
49	CG57721-01	115	116	Prestin
50	CG57787_01	117	118	Sulfate Transporter
51	CG57785_01	119	120	Sulfate Transporter
52	CG57748-01	121	122	N-acetylgalactosaminyltransferase
53	CG57693-01	123	124	Protein Kinase
54	CG57707-01	125	126	Leucine-rich glioma-inactivated protein precursor
55	CG57306-01	127	128	Anion exchanger
56	CG57348_01	129	130	PR_SET domain protein
57	CG57650-01	131	132	NONMUSCLE MYOSIN HEAVY CHAIN B
58	CG57766-01	133	134	Plasma retinol binding protein
59	CG57566-01	135	136	HIV-1 inducer of short transcripts binding protein
60A	CG57574-01	137	138	Beta tectorin
60B	CG57574-02	139	140	Beta tectorin
60C	CG57574-03	141	142	Beta tectorin
61	CG57505-01	143	144	KIAA1125
62A	CG57473-01	145	146	Zinc-finger protein BOP
62B	CG57473-02	147	148	Zinc-finger protein BOP
63	CG57777_01	149	150	Hypothetical secreted protein
64	CG57779_01	151	152	Hypothetical secreted protein
65	CG57781_01	153	154	Hypothetical secreted protein
66	G57783_01	155	156	Hypothetical secreted protein
67A	CG57823-01	157	158	ACYLTRANSFERASE
67B	CG57823-02	159	160	ACYLTRANSFERASE
68	CG57801-01	161	162	guanine nucleotide exchange factor
69A	CG57719-01	163	164	Aspartate Aminotransferase
69B	CG57719-02	165	166	Aspartate Aminotransferase
70	CG57462-01	167	168	KIAA1337
71	CG57584-01	169	170	ZONA PELLUCIDA GLYCOPROTEIN 1 PRECURSOR
72	CG56761-01	171	172	Ankyrin repeat containing protein
73	CG57313_01	173	174	GPCR
74	CG57315_01	175	176	GPCR
75	CG57317_01	177	178	GPCR
76	CG57321_01	179	180	GPCR
77	CG57419_01	181	182	GPCR
78	CG57425-01	183	184	GPCR
79	CG57753-01	185	186	GPCR
80	CG56766-01	187	188	GPCR
81A	CG57847-01	189	190	GPCR
81B	CG57847-02	191	192	GPCR
82	CG57845-01	193	194	GPCR
83	CG57843-01	195	196	GPCR
84A	CG57841-01	197	198	GPCR
84B	CG57841-02	199	200	GPCR
85	CG57839-01	201	202	GPCR
86	CG57837-01	203	204	GPCR
87	CG56763-01	205	206	GPCR
88	CG56753-01	207	208	GPCR
89	CG57670-01	209	210	GPCR

CTGGCGGGTGCTGCGGACTCCACCTGGGCTGTCCACCTAACACCCCGGGGACCGCTGCCCC
CAGCCTGGACACTGTGCAGCGGGTCCAGGTGACTATGGGACCCGAGGAGGTGCTGATCCAG
GCTGGAAATGTGTCTGTGAAGGGGCAGCTGGTACCTGAAGGGCAGTCTTGGCTGCTCCAC
GGGCTGAGCCTGCAATGGCTGGGGGACTGGCTGGTGTCTGTGAGGAGGCTGGGGGTGCTG
GTGCGGCTGGACGAGTGGCTCCATCTCCATCTCTGTGAGACACGAGCTCTGGGGACAG
ACACAAGGCCTCTGTGGGCTCTACAATGGCTGGCCAGGAGATCTTATGGAGCCAGGC
GGAGGGCTGGCCATGTTAGCAGCCACCTTTGGAAATTCTCGAGGCTCCCTGGCTCGGAG
GTTTCCCCCGCTGAGTACCACGAGGCCTGTCTCTTTGCCTACTGCGCAGGGGCCATGGCA
GGCAGTGGGCAAGAGGGGGCGGCAGCAGGCTGTTTGTGCCACCTTTGCCAGCTATGTCCAG
GCCTGTGCCAGGCGGCACATCCACATTGCGTGGAGGAAGCCTGGCTTCTGCGAGCGCCTG
TGCCCCGGGGGCCAGCTCTACTCCGACTGCGTCTCCCTCTGCCACCCAGCTGCGAGGCG
GTGGGTGAGGAGAGGAGGTCTCTGCAGGGAAGATGTGTGAGTGGCTGTGAGTGGCCCG
CGAGGCCTCTTCTGGAATTGGCACCTCTGTGTGCTGCTGCGCCACTGCCCTGCTACTAC
TGCCGCCAGCGCTATGTACCCGGTGACACCGTGCGCCAGCTGTGTAACCCCTGCGTGTGC
AGGGATGGCCGCTGGCACTGTGCCCAGGCACCTGTGCCCGCCGAGTGTGCAGTGGGTGGG
GACGGGCACTACCTCACCTTCGATGGGCGGAGCTACTCTTCTGGGGTGGTCAAGGTTGC
CGCTACAGCCTGGTGCAGCCCTATCCTTCCACGTGCCCCACCCCCACCATCAGACCCCT
GTTCCAGGAGCTGTGCTGGTCAATGGGCAGGATGTGGGCTTGCCCTGGATTGGCGCTGAG
GGCCTCAGTGTGCGCGCAGCTTCTCTGCTTCTGCTGCTGCGCTGGCCTGGGGCCAG
GTGCTCTGGGACTGTGTGACCTGTAGCCTACATACCTGGACCCCCGCCATGCCAC
CAGGTGCAGGCTGTGTGTGGCACCTTACCCAGAACCAGCAGGACGACTTCTGACACCA
GCCGGAGATGTGAAACTAGCATTGCTGCCTTTGCTAGCAAGTTCCAGGTGGCCGGCAAG
GGAAGATGCCCCCTGTGAGGACAGTGCCTGCTGTCTCCCTGCACACCCACTCCCAGCGC
CACGCCTTCGCAGAGGCGGCCTGTGCCATCCTGCACAGCTCTGTCTTCCAGGAATGCCAC
AGGCTGGTGGACAAAGAGCCATTCTATCTGCGCTGCCCTGGCAGCCGTGTGTGGCTGTGAT
CCCCGCAGTGACTGCGCTGTGCCCCGTGCTGTCTGCCATGCGCGCTGCTGCCAGGAA
GGTGCTTCCGACTCTCCCTGGAGGAACGACCCCTGCTGCCCTGTTATGTGTCTGGTGGCCAG
GAGTACCGAGAGTGTGCCCCAGCATGCGGTCAACACTGCGGGAAACAGAGGACTGTGGA
GAGCTGGGCAGCTGTGTGGCTGGTTGTAAGTGTCTCTGGGGCTGCTGTGGGACCCTGAG
GGCCAGTGTGTGCCCCCAGCTTGTGCCCTGCCAGCTCGGAGCCCGTGCCTATGCCCT
GGCAGTGCCACCATGAAGGAGTGCAACCGCTGGGAGCTTGTCTATGCCCTGGTGCCTGT
CTCCTACCTGTGACAGCCCCAGGCCCAATCACTCCTGCCCTGCGAGCAGTACTGATGGC
TGTGTCTCTCCACAGGACGGTGTCTGTGACGAGCGCTGTGTGCTCTGACCTCTGT
CCTGCCGTACAGTGGGCAAGTGGTACCTGCCCAACGCCACCATCCAGGAAGACTGCAAC
GTTTGGCTGTGCCCCGGGCGGCAGTGGCACTGCACAGGCCAGCGGCGAGTGGGCGGTGC
CAGGCATCAGGCGCCCCCACTATGTGACATTTGACGGACTGGCCTTCACTATCCTGGG
GCCTGCGAGTATCTGCTGGTGCAGAGGCCAGTGGCTATTACAGTCTCTGCCCAGAAC
CTGCCCTGTGGGGCCAGCGGTCTCACCTGCACCAAAGCGCTGGCCGTGCGTCTGGAGGGC
ACTGTTGTGCACATGCTCAGAGGGACTCGGGTCTGGTGCAACTGTCCCCTCAGTTCCGT
GGTCGCTGGCTGGGCTGTGGTGACTTTGATGGAGATGCCAGTAATGATCTGCGGAGC
GGCCAGGCGCTCTGGAGCCACAGCTGAACCTGGCTGCCACTCTGGCGCCTCAGCCCC
CTCTGCCCTGAGCCAGGAGACCTGCCACACCCCTGCACGATGAACACACACCGGGCTGGT
TGGGCTCGGGCCCGCTGTGGGGCGCTGCTGCAGCCGCTCTTACATTATGCCACGCGGAG
GTCCCCCGCAGCAGCACTATGAGTGGTGCCTGTATGACGCTGCGGCTGCGACTCGGGG
GGTGACTGTGAGTGCCTCTGCTCGGCCATTGCCACCTATGCAGATGAGTGTGCCGGCAT
GGGCACCACTGCGCTGGCGTAGCCAGGAGCTCTGCTGTCTACACAGACACCCCTGTGCC
CTCCATTGGTGGTCACTTTGGCCAGCCAGCTGGTGTGTTGATCTTTTGGCCTTGTGC
CTCAGTGATCCAAAGCTCTCTCCTCTCCACCCAGCGCTGCAGTGTGAAGGGGAGAGTA
TATGAGGCCTGTGGCCCCACGTGTCCCCCACCTGCCATGAGCAGCATCCTGAGCCCCGG
TGGCACTGCCAGGTGGTGGCCTGTGTGGAGGGCTGCTTCTGCCCGAGGGGACTCTGCTG
CACGGACGCTGTGTGCTCAGCACGTGCCAGGAAGTCAATGGCATTGTGGGGGTGACGGT
GGCCACTGTGAGGAGCTTGTGCCTGCCTGTGACAGAGGGAGAGGCCCTGTGCCAAGAGAAT
GGGCACCTGTGTGCCCATGGGTGGCTTTGTGACAACACAGGACGATGTGGCGATGGCTCT
GATGAGGAGGGTGAGTGTCTTTGCCATGCGTGGAAAGCAGCAGGGTTGGTCACTCTTGC
ACATGTTGTGCGCGCCCCAGGCTGTGGGGAGGGGCAGATGACTTGCAGTCCGGCCACTGC
CTGCCCTTGGCCCTGCTCTGTGACCGCCAGGATGACTGTGGAGATGGCACGGATGAGCCG
AGCTATCCGTGCCCCACGGGCTTGCTGGCCTGTGCCGATGGACGCTGCCTGCCGCCGGCC
CTGCTCTGCGATGGGCATCCTGACTGTCTGGATGCCGCCGACGAGGAGTCTGTCTGGGG
CAGGTGACCTGCGTCCCCGGGGAGGTGTCTGTGTTGATGGCACCTGCCTGGGGGCCATC
CAGCTGTGTGACGGAGTCTGGGACTGCCAGATGGAGCCGATGAGGGGGCCGGACACTGC
CCCCACTCTCTCTGCCACACCTCTGCCAGCACCTTGCCCTGGCCCCCTCCGACGCTCC
CTGGACACTGCGTCAAGTCCCCCTGGCCAGCGCCAGCCCTGCGCCACCCTGCGGCCCTTC

GAGTTTCGGTGC CGGCAGCGGCGAGTG CACCCGCGGGGCTGGCGCTGCGAC CAGGAGGAA
 GACTGCGCCGACGGCAGCGAGCGCGGCTGCGGAGGGCCCTGCGCGCCGAC CACGCG
 CCCTGCGCCCGCGGCCCTCACTGCGTGTCCCCCGAGCAGCTGTGCGACGGCGTGCGGCAG
 TGTCCCGACGGCTCGGACGAGGGCCCCGACGCTGCGTTGAGGCTCCCGCGCCCCCGGCC
 ATGCGCGGCCCCCTGGCCAAGCCGGCGGGCCACCTCTTCCCGAGCGCCATCCCCACCT
 TCGCCTCCTGAGGCACAGGGAGAGGGCAGGAAGGGACAGGAGCGGAGCAGGACACATCTC
 ACAGTGCCCGCAGGCTCCACCCAGCTGCCTCTGTGCCCTGGCCTCTTTCCCTGTGGTGTG
 GCTCCGGGGCTGTGCCTGACCCCTGAGCAGCTCTGTGATGGGATCCCAGACTGTCCCCAG
 GCGGAGGACGAGCTGGACTGCGGGGGGCTGCCAGCCCTGGGAGGCCCCAACAGGACAGGG
 CTTCCCTGCCCAGAATACACCTGCCCCAATGGCACCTGCATAGGCTTCCAGCTGGTGTGT
 GATGGGCAGCCTGACTGTGGAAGGCCAGGGCAGGTGGGGCCCTCCCCAGAAGAGCAGGGT
 TGTGGGGCCTGGGGCCCTGGAGCCCATGGGGGCCCTGCAGCCGGACGTGTGGGGCCCTGG
 GGCCAGGGCCGGAGCCGCGCTGCTCCCCACTCGGCCCTCCTGGTGCTACAGAACTGCCCA
 GGGCCTGAGCACCAGTCTCAGGCCTGCTTACGGCAGCCTGCCAGTGACGGTGAATGG
 AGCACCTGGTCCCCCTGGTCTGTGTGCTCTGAGCCGTGCAGGGGCACCATGACGCGGCAA
 CGGCAGTGCCACTCACCCAGAATGGGGGCCGACCTGTGCTGCACTGCCCGGAGGCCCTG
 CACAGCACCCGCCAGACCAAGCCTTGCCCTCAGGACGGCTGCCCAATGCCACTTGTCTCT
 GGGGAGCTGATGTTCCAGCCCTGTGCCCTTGCCCACTGACCTGTGATGACATCTCTGGC
 CAGGTACGTGCCACCTGATTGGCCCTGCGGCAGCCCGGGCTGCTGGTGGCCAGAAGGG
 CAGGTGCTGGGCAGCGAGGGGTGGTGTGTGTGGCCCCGGCAGTGCCCTGCCTGGTGGAC
 GGTGCCCCGCTACTGGCCTGGGCAACGCATCAAGGCCGACTGCCAGCTCTGCATCTGCCAA
 GACGGACGGCCCCGACGCTGCCGACTCAACCCGACTGCGCTGGTGAGGCCCTTCCCTCG
 GGGTCCCTAGTCTCTCCCTGGACCGCCAGCTGCACATCCACCACCTCCTTCAGGCTCT
 GACTGTTGGCCCTCCCTCAGTGGACTGTGGCTGGTCCCTCCTGGTCAACCCTGGGCCAAGTG
 CCTGGGCCCTGTGGAAGCCAGAGCATCCAGTGGTCCCTTCCGGAGCTCCAACAACCCCG
 CCCCTCCGGCCGAGGTGCGCCAGTGCCGTGGCATCCACCGCAAGGCACGCAGACGGAGCCC
 TGTGAGGGGTGTGAGCATCAGGGCCAGGTCCACCGTGTGCGGGGAACGCTGGCATGGGGGC
 CCCTGCAGGGTGTGCCAGTGTCTGCACAACCTCACCGCACACTGCTCACCTACTGCCCG
 CTCGGCAGCTGCCCCAGGGCTGGGTCTTGGTGGAGGGGACGGGAGAATCATGCTGCCAC
 TGTGCCCTACCTGGAGAGAACCAGACGGTCCAGCCCATGGCCACTCCTGCCGACGCTCCG
 GCTCCAGTCCCCAGATCAGATTCCCTTTGGCCACTTACATTCTGCCTCCGTCAGGAGGC
 TCCTGCGCCCTCTGTCTCTCCCTACTCCAGCCTGTCTCTCTCTTCTGCACCCAGACCCC
 TGTATTCTCCCTTGGGGCTGGCCGACTGGCTGAGGGGAGTCTGCATGCATCGTCCCAG
 CAGCTGGAACACCCACCCAGGCTGCCCTCCTGGGGGCTCCACCCAGGGGGCCAGCCCT
 CAGGGATGGCACGCTGGAGGGGATGCTTATGCCAAGTGGCACACTCGGCCCATTACCTG
 CAGCTGGACCTGCTTCAGCCTCGGAACCTCACTGGCATCCTAGTGCCGGAGACTGGCTCC
 TCCAACGCATATGCCAGCAGCTTCTCACTCCAGTTCAGCAGCAATGGTCTACACTGGCAT
 GACTATCGTGACCTCCTGCCTGGCATCTTGCCCTGCCCAAGGTATCACCCGCCCAAGGC
 CGATGGGGCCAGCAGCCCAACCATGCCCTTTTGTGGGTTCATAGTCTTTGTCCCCAAGGG
 CTTTCCAGTGTCCCCGAGGGGCATGGCTGCATTGATGCTTGTGAATACCTGCTTTTC
 CCTAGAACTGGGATGACCTGGACCTGCCGTATGGACTTTGCGCCGCATGGTGCAGGCG
 AGGTTTGTGAGGGTGTGGCCCCACGATGTCCACCACAGCGATGTCCCCCTGCAGGTGGAG
 CTGCTGGGCTGCGAGCCAGGGGTGGACTCCGCTGTGCCAGTGGTGAGTGTGTCTTGAGA
 GGGGGCCCTTGTGACGGTGTCTGGAAGTGGCAGGATGGCTCGGATGAGGAGGGCTGTGTG
 TTGCTGCCTGAGGGCACTGGCAGGTATAC'TGTGGCCGGCCGTGCAGCTCACGCCCTTGGC
 CTGGCCTTTGAGGGGACAGCCATGTGGGAGGGGGCCCGGCACTGCCTTCACCCCCAAGGTG
 CCCAGACCTGCATGCTGAGGAGCTGCAGCCGGGGCCTGGCAGAGACTGAGCACTGGCCC
 CCTGGGCAGGAATCCCCCACGTCCCCGACAGAAGCCTGGGATACCTCTCGAGGGCCCCC
 ACCTTCCTTTTCTGGGAAGGGGAGCTGGGCAAGCCTCACCTCCCTCTACCTACCCACA
 GAGACAAGGCCCGTGAGTCTGGCCAGCCTCCGGGGTGCTCACCATGGGGAATCTGTG
 CAGATGGTGACCACCAACCCCATACCCAGATGGAGGCCAGGACCTGCCACCAGGTATG
 GCAGCTGTGACGGTGGTGCCCCACACCTGTGACTCCAGCGACCCCTGCTGGTAAAGATG
 CCAAGCCCTTTACCTGTGTGCAAGGCCCTCTGCCAGAGCGTCGCCCCAGGACCCCTTCCCA
 CCTGTGACAGTGTGGCCCCGGCCAGACGCCCTGTGAGGTGCTGGGCTGCGTGGAACAGGCG
 CAGGTGTGTGATGGCAGGGAGGATTGCCTCGACGGCTCCGACGAGAGGCACTGCGGTGAG
 CTGCTGGAGGGCTGTGCTGTCTGTGAGCGGCGCTGTGACCTGCGGCCTGACTGCCAGGATGGC
 TCGGACGAGGATGGCTGTGTGGAAGTGGCTGTGCCCCCTGGTCTGTCTGGAGCAGCTGC
 AGCCGACAGCTGTGGCCTGGGCCTCACCTTCCAGCGCCAGGAGCTGCTGCGGCCTCCTCTG
 CCAGGGGGCAGCTGCCCCGCTGACCGGTTCCGAAGCCAGTCCCTGCTTTGTGACGGCCTGC
 CCAGTGGCTGGGGCATGGGCCATGTGGGAGGCCCTGGGGACCCCTGCAGCGTCTCTGCGGG
 GGTGGCCATCAGAGTCGCCAGAGAAGCTGTGTGGACCCCCACCCAAGATGGCGGTGCC

CCCTGCCCCCGGGGCCCTCCCAAGAGAGAGGGACACCTTGCGGCTTTGCAGCCCTGCTCAGGTGGC
ACAGGTAAGGGGGTGCTGGGCTGGGGACACGGAGGAGCACTGTGGGCACGGGGCGATTG
GGCCTCCACAGCACTAGGCTCACCTGGTGCCCATCCCCACCACTGCGAGCTGGGCGG
TGTGTATGTGAGTGCCGATCTGTGCCAGAAGGGGCTGGTGCCCCCATGCCACCCCTCCTG
CCTGGATCCCAAGGGCAACAGAAGCTGCAGTGGGCATTTGTGTGGAAGCCTCCCCCTCCTG
CTGTGCCCTTTGGGGCTGAGTGCCCTCTTCCACATACTCCAGGATGCCGCTGCCCCCG
GGGCTCCTTCTGTCATGACACTCGCTGCCCTGCCCTCTCTGAGTGCCCCCTGCCCTGGTGGGC
GAAGAGCTGAAGTGGCCAGGGGTGTCTTCTCCTGGGCAACTGCAGCCAATGCGTGTGT
GAGAAGGGGGAGTTGCTGTGCCAACCAGGGGGTGCCCCCTGCCCTGCGGCTGGTCAGCC
TGGTCTCTCTGGGCTCCCTGCGACCGCTCCTGTGGCTCTGGAGTGAGGGCCAGGTTACAG
TCTCCCTCCAACCCCTCCGGCAGCCTGGGGGGGTGCCCCGTGTGAAGGTGACCGGCAGGAA
CTGCAGGGCTGCCACAGTGTGTGGGACAGGTATAGCCGGGAGCCTGGGTGCAGGAGTCA
CCCCCTCTCTCTCCCAATTCTGTACCTTGAGAACACATGGGATGGGACCCACTGACCAC
TCTACGTGGGGGATTGAGGTGTTTCGGCTGGACGCCCTGGACTTCTTGGTCTCTGCTCC
CAAAGCTGCCCTTGCCCCGGGAGGGGGCCCTGGCTGGCGCAGTCGTTCCCGACTCTGCCCC
AGCCCTGGGGATTATCTCTGCCCAGGAGATGCCACCCAGGAGGAGCCCTGCAGCCCCCT
ATAGAGTGTACGGGCTTCTGCGCCCCCGGCTGCACCTGCCCCCTGGTCTTTCTCTGCAC
AATGCTAGCTGCCTGCCCCGACGCCAGTGCACCCCTGCCAGCTGCACGGGCAGCTCTATGCA
TCAGGAGCAATGGCTCGCTGGACTTCTGCCAACAACCTGCACCTGTGTCTCTGGTAAAGATG
GCATGCACCTCGGAGCGCTGCCCAGTGGCCTGTGGTGGAGTCCCTGGACCCCTGTGGAGT
CTCTGTAGCTGCAGCTGCAACGTGGGCATTCGGCGCCGCTTCCGGGCAGGCACTGCACCC
CCAGCTGCCCTTTGGGGGTGCTGAGTGCCAAGGCCCCACCATGGAGGCTGAATTTCTGCAGC
CTGCGGCCATGTCCAGGGCCAGTGCCTGGCATGTGTCCAGGGACAAGCAGTGGCTGGAC
TGTGCCCAGGGCCCTGCCTCTTGTGTCAGAGCTCAGCGCCCCAAGAGGGACTAACCCAGACC
TGCCACCCCTGGCTGCCACTGCCCTCTGGGATGCTTCTGTGTTGAGCCACGTGGTCCAC
CCTGGACCCCTTGGAGCCAGTGTTCAGCCTCCTGTGGCCCTGCCCGGTGCCATCGGACCC
GGTCTGTGCCAGGAGCTGGGGATGGGGTCCATGGGGGCCCTGGTCCCACTGTAGCCCG
AGCTGTGGGGGAGGCCTGCGGAGCCGACCCGGGCCTGTGACCAGCCCCACCCAGGGC
CTGGGGGATTACTGCGAGGGGCCACGGGCACAGGGGGAGGTCTGCCAGGCTCTGCCCTGC
CCAGTGACCAACTGCACCTGCCATTGAAGGGGCCAGTATAGCCCTGTGGCCCTCCGTGC
CCTCGCTCCTGTGATGACCTAGTGCAC'TGCGTGTGGCGCTGCCAGCCTGGCTGCTACTGC
CCACCAGGCCAGGTACTGAGTTCCAACGGGGCCATCTGCGTGCAGCCGGGTCACTGCAGC
TGCTTGACCTGCTGACCGGGCAGCGGACCATCGGGTGTCTGGCTGGCAAGGCTGCAG
GGCTGCAACCTACTGCACCTGCTGGAGGGGAGGCTGAACTGCACAGACCTGCCCTGCCCA
GACTGCGGGGGTGGCCAGAGTCTGCATCCCTGTGGGCAGCCCTGCCCCCGCTCCTGCCAG
GACCTGTCCCCTGGGAGTGTGTGCCAGCCAGGCTCTGTGGGCTGCCAGCCACTTGTGGG
TGCCCCCTGGGGCAGCTCTCCAGGACGGGCTGTGCGTGCCCCAGCCCACTGCCGCTGC
CAGTACCAGCCTGGAGCCATGGCCCCCTCCTTTGTCCCCAGCACTGTGTGGCAGGCATT
CTGCAATGCCAGAGGTGCTGACTGCCCGGACCTGGGGTGTGGAGCTCTTGGGGCCCT
TGGGAAGACTGCAGTGTTCGTGTGGGGCGGGGAGCAGCTGCGCTCCCGCGCTGTGCT
CGTCTCCTGCCCAGGGCCTGCCCGCCAGAGCCGCACATGCAGCACACAGGTCTGCAGA
GAGGCAGGCTGCCCGGCTGGCCGCCTGTACCGTGAATGCCAGCCCGGCGAGGGATGCCCC
TTCTCCTGCGCCACGTACGCAGCAGGTGGGCTGCTTCTCTGAGGGCTGCGAGGAGGGC
TGCCACTGCCCGAGGGCACCTTCAGCACCGCCTGGCCTGTGTGCAGGAGTGCCCTTGT
GTGCTGACAGCCTGGCTGCTGCAGGAGCTGGGAGCCACCATAGGTGACCTGGTCAAGCC
CTCGGCTGCTGAGATGAGCTGGACTCAGGCCAGACATTCGTACAAGCTGTGGCACTGC
TCGTGTGCACACGGGAAGCTGTCTTGCTCCTTGAGCAGTGTCTGAGGCGTGAAGTGGT
TTCGGTCCCTGGAGCCCGTGGGGCCCGTGCTCCCGCTCCTGTGGAGGGCTGGGCACCCGT
ACCCGAGCCGCCAGTGTGTGCTACCATGCCCACCTCAGTGAGCTGCCCGTGTGCCCT
GGCCCTGGCTGTGGGGCTGGGAAGTGTCTTGACCTCCTGGGGCCCGTGGGAAGCTTGC
TCCCGCAGCTGCGGAGTGGGCCAGCAGCGCCGCTGCGGGCATAACCGTCCCCCTGGGCCC
GGCGGGCACTGGTGCCCCAACATCTTACTGCCTACCAAGAGGCGCCGCTTCTGCAACCTG
CGAGCTGCTCCAGAGCCGGCTGCCCAGCAGGATGGAGGTGGTCACTGTGGTCCCAACCG
TGCCCCCGCGCTGCTGCAGACCTCCAGAGGGGAATTGTGTGTGTCAGGACGACCAAGTCTGC
CAGAAGGGCTGCCGCTGCCCAAAGGGGTCCCTGGAGCAGGATGGTGGCTGCGTGCCAATT
GGGCACTGTGACTGCACCGATGCCCAGGGCCACAGCTGGGCCCCGGGGAGCCAGCACCAG
GATGCTGCAACAAGTGTCTATGCCAAGCTGGGCAGCTCTCCTGCACGGCTCAGCCCTGC
CCGCTTCCACCCACTGTGCTTGGAGCCACTGGTCCGCTGGAGTCCCTGCAGCCACTCA
TGCGGGCCCCAGAGGGCAGAGAGCCGCTTCCGCTGCGGCGCGGGCTGGCCCTCTCGCTCT
GGTCTCTCCCTGCTGATGGCCAAGGCCGACCCCACTGCACAACAGCACTTCTCCAC
CTGGACACCCAGGGCTGCTACTCAGGGCCCTGCCAGACTCATGCCAGTGGAGTCTGTGG
GGGCCATGGAGCCCTGCCAGGTGCCCTGCAGTGGGGGGTTAGGCTACGCTGGAGAGAG

```

GCAGAGGCCCTCTGTGGAGGAGGCTGCCGGGAGCCATGGGCTCAAGACAGAAAGCTGCAA
CGGAGGGCCCTGCCAGCACCTGTGTCAACGAGTCCCTGGTGTGCCACACCAGGAGTGT
CCAGTCCCTTGGGCTTGGTCAGCCTGGAGCAGTTGCTCGGCCCCCTGTGGTGGGGGCACT
ATGGAGCGACATCGGACTTGTGAGGGGGGTCTGGGGTGGCACCATGCCAGGCCCAGGAC
ACAGAGCAACGGCAGGAGTGTAACCTGCAGCCCTGCCCTGAGTGCCCCCTGGCCAGGTG
CTTAGTGCCCTGTGCCACCTCATGCCCGTGCCTCTGCTGGCATCTGCAGCCTGGTGCCATC
TGTGTGCAGGAGCCCTGCCAGCCTGGCTGTGGCTGCCCTGGAGGGCAGCATTCTCTGCCC
TGGGGCCTCACCCCTGACCCTGGAAGAGCAGGCCCAGGAGCTGCCCCAGGGACTGTGCTC
ACCCGGAAC TGACCCGCTGTGTCTGCCACGGTGGAGCCTTCAGCTGCTCCCTCGTTGAC
TGTCAGGGTGAGATAGTGCCCCCTGGGGAAACGTGGCAGCAGGTGGCCCCGGGGGAGCTG
GGGCTCTGCGAGCAGACGTGCCTGGAGATGAACGCCACAAAGACCCAGAGTAACTGCAGT
TCAGCTCGAGCCTCGGGCTGCGTGTGCCAGCCCGGGCACTTCCGCAGCCAGGCAGGCCCC
TGCGTCCCCGAAGACCACTGCGAGTGCTGGCACCTTGGGCGTCCCCACCTGGTGAGACAC
CGAACCCCTCTGCTACCACTCACCCATTCTGACCCCAAGCCTCCCATCTGTCTGTAA

```

In a search of public sequence databases, the NOV1 nucleic acid sequence, located on chromosome 8 has 606 of 779 bases (77%) identical to a gb:GENBANK-ID:BTSCOSPON|acc:X93922.1 mRNA from *Bos taurus* (*B.taurus* mRNA for SCO-spondin protein). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

In all BLAST alignments herein, the “E-value” or “Expect” value is a numeric indication of the probability that the aligned sequences could have achieved their similarity to the BLAST query sequence by chance alone, within the database that was searched. For example, the probability that the subject (“Sbjct”) retrieved from the NOV1 BLAST analysis, *e.g.*, *B.taurus* mRNA for SCO-spondin protein, matched the Query NOV1 sequence purely by chance is $2.4e-152$. The Expect value (E) is a parameter that describes the number of hits one can “expect” to see just by chance when searching a database of a particular size. It decreases exponentially with the Score (S) that is assigned to a match between two sequences. Essentially, the E value describes the random background noise that exists for matches between sequences.

The Expect value is used as a convenient way to create a significance threshold for reporting results. The default value used for blasting is typically set to 0.0001. In BLAST 2.0, the Expect value is also used instead of the P value (probability) to report the significance of matches. For example, an E value of one assigned to a hit can be interpreted as meaning that in a database of the current size one might expect to see one match with a similar score simply by chance. An E value of zero means that one would not expect to see any matches with a similar score simply by chance. See, *e.g.*, <http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/>. Occasionally, a string of X’s or N’s will result from a BLAST search. This is a result of automatic filtering of the query for low-complexity sequence that is performed to prevent artifactual hits. The filter substitutes any low-complexity sequence that it finds with

the letter "N" in nucleotide sequence (e.g., "NNNNNNNNNNNNNN") or the letter "X" in protein sequences (e.g., "XXXXXXXXXX"). Low-complexity regions can result in high scores that reflect compositional bias rather than significant position-by-position alignment. (Wootton and Federhen, *Methods Enzymol* 266:554-571, 1996).

- 5 The disclosed NOV1 polypeptide (SEQ ID NO:2) encoded by SEQ ID NO:1 has 4219 amino acid residues and is presented in Table 1B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV1 has a signal peptide and is likely to be localized extracellularly with a certainty of 0.5087. The most likely cleavage site for NOV1 is between positions 17 and 18.

Table 1B. Encoded NOV1 protein sequence (SEQ ID NO:2).
MLLPALLFGMAWALADGRWCEWTETIRVEEEVAPRQEDLVPCASLDHYSRLGWRLDLPWS GRSGLTRSPAPGLCPIYKPPETRPAAKWNRTVRTCCPGWGGAHCTEALAKASPEGHCFAMW QCQLQAGSANASAGSLEECCARPWGRSVDGSSQACRSCSSRHLPGSASSPALLQPLAGA VGQLWSQHRPSATCASWSGFHYRTFDGRHYHFLGRCTYLLAGAADSTWAVHLTPGDRCP QPGHCQRVQVTMGPEEVLIQAGNVSVKQQLVPEGQSWLLHGLSLQWLGDWLVLSGGLGVV VRLDRTGSSISISVDHELWGQTQGLCGLYNGWPEDDFMEPGGGLAMLAATFGNSWRPLPGSE VSPAIEYHEACLFAYCAGAMAGSGQEGRQAVCATFASVYQACARRHIHIRWRKPGFCERL CPGGQLYSDCVSLCPPSCAEAVGQGEESCREECVSGCECPRGLFWNGTLCVPAHPCPCYY CRQRYVPGDTRVQLCNPCVCRDGRWHCAQALCPAECVGGDGHYLTDFDGRSYSFVGGQGC RYSLVQPPYSTCPTPTIRPPVPGAVLVNGQDVGLPWIGAEGLSVRRASSAFLLLRWPGAQ VLWGLSDPVAYITLDPRAHQVQGLCGTFTQNNQDDFLTTPAGDVETSIAAFASKFQVAGK GRCPSEDSALLSPCTTHSQRHAFEAACAIIHSSVFQECHRLVDKEPFYLRCLAAVCGCD PGSDCLCPVLSAYARRCAQEGASPPWRNQTLCVPMCPGGQYRECAPACGQHCGKPEDCG ELGSCVAGCNCPLGLLWDPEGQCVPPSLCPCQLGARRYAPGSATMKECNRWELVYAPGAC LLTCDSPSANHSCPAGSTDGCVCPPGTVLLDERCVPPDLPCRHSQGWYLPNATIQEDCN VCVCRGRQWHCTGQRRSGRCQASGAPHYVTFDGLAFTYPGACEYLLVREASGLFTVSAQN LPCGASGLTCTKALAVRLEGTVVHMLRGRTRVLVQLSPQFRGRVAGLCGDFDGDASNDLRS RQGVLEPTAELAAHSWRLSPLCPEPGDLPHPCMTMNTHRAGWARARCGALLQPLFTLCHAE VPPQQHYEWCLYDACGCDSGGDCECLCSAIATYADECARHGHVVRWSQELCCLHQTPCA LHGGHLGQPAWCGCILLPLCLSDPRLSPLHPALQCEGGQVYEACGPTCPPTCHEQHPEPG WHCQVACVEGCFCEPGLLHGRVCLSTCQEGQWHCGDGGHCEELVPACAEGEALCQEN GHCVPHWGLCDNQDDCGDGSDEEGECLCPCVEATGLVSPCTCCAAPGCGEGQMTCSGHC LPLALLCDRQDDCGDGTDEPSYPCPQGLLACADGRCLPPALLCDGHPDCLDAADEESCLG QVTCVPGEVSCVDGTCLGAIQLCDGVWDCPDGADEGPGHCPLPSLPTPPASTLPGSPGS LDTASSPLASAPAPPCGPFEFRCGSGECTPRGWRCDQEEDCADGSDERGC GGPCAPHHA PCARGPHCVSPEQLCDGVRCPDGSDGPDACVEAPAPPAMRGPPGQAGGPTSSRAPSP SPPEAQGEGRKGQERSRTHLTVPAGSTQLPLCPGLFPCGVAPGLCLTPEQLCDGIPDCPQ GEDELDCGGLPALGGPNRTGLPCPEYTCPNGTCIGFQLVCDGQPDGCRPGQVGPSPPEQG CGAWGPWSPWGPCSRTCGPWGQGRSRRCSPLGLLVLQNCPEHQSQACFTAACPVDGEW STWSPWVCSEPCRGTMTRQRQCHSPQNGGRTCAALPGGLHSTRQTKPCPDGCPNATCS GELMFQPCAPCPLTCDDISGQVTCPPDWPCGSPGCWCPEGQVLGSEGWCVWPRQCPCLVD GARYWPGQRIKADCQLCICQDGRPRRCRLNPDCAEALPSGSLVLSLDRPAAHPPPPSGS DCWPSLSGLWLVLLVTLGQVPGPLWKPEHPVVLPELQQPPPLRPRSPVPWHPPQGTQTEP CEGCEHQVHRVGERWHGGPCRVCQCLHNLTAHCSPYCPLGSCPGQVWLVEGTGESCC CALPGENQTVQPMATPAAAPAPSPQIRFPLATYILPPSGGSCRPLSSPTPACLSLLHPDP CYSPLGLAGLAEGSLHASSQQLHPTQAALLGAPTQGPSQGWHAAGDAYAKWHTRPHYL QLDLLQPRNLTGILVPEGTSSNAYASSFSLQFSSNGLHWHYRDLLPGILPLPKVSPAQG RWGQQPTMPFCGFHSLCPQGPSSVPEGHGLHSMLEVEYLLFPRNWDLDPAVWTFGRMVQA RFVRVWPHDVHSDVPLQVELLGEPEGVGLRCASGECVLRGGPCDGVLDCEDEGSDEEGCV LLPEGTGRYTVAGRAAHALGLAFEGTAMWEGPGTAFTPKVPRPCMLRSCSRGLAETEHWP PGQESPTSPTEAWDTLSRAPTFLSWEGELGKPHLPLPTETRPVSPGPASGVRPHHGESV QMVTTTTPIQMEARTLPPGMAAVTVVPPHPVTPATPAGKMPSPSTCVQGLCQSVAPGPF PVQCGPGQTPCEVLGCVEQAQVCDGREDCLDGSDEHRCGELLEGLSCGALCSPSOLSCG

gi 1711548 sp P98167	SSPO BOVIN	SCO-spondin, bovine	867	582/1035 (56%)	635/1035 (61%)	0.0
gi 5823595 emb CAB537	60.1 (AJ132107)	SCO-spondin, bovine	564	360/494 (72%)	395/494 (79%)	e-173

The presence of identifiable domains in NOV1, as well as all other NOVX proteins, was determined by searches using software algorithms such as PROSITE, DOMAIN, Blocks, Pfam, ProDomain, and Prints, and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (<http://www.ebi.ac.uk/interpro>). DOMAIN results for NOV1 as disclosed in Tables 1D-H, were collected from the Conserved Domain Database (CDD) with Reverse Position Specific BLAST analyses. This BLAST analysis software samples domains found in the Smart and Pfam collections. For Tables 1G-1O and all successive DOMAIN sequence alignments, fully conserved single residues are indicated by black shading or by the sign (!) and “strong” semi-conserved residues are indicated by grey shading or by the sign (+). The “strong” group of conserved amino acid residues may be any one of the following groups of amino acids: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW.

Tables 1D-H list the domain descriptions from DOMAIN analysis results for NOV1. This indicates that the NOV1 sequence has properties similar to those of other proteins known to contain this domain.

Table 1D. Domain Analysis of NOV1

gnl|Pfam|pfam00754, F5_F8_type_C, F5/8 type C domain. This domain is also known as the discoidin (DS) domain family.

CD-Length = 145 residues, 100.0% aligned

Score = 115 bits (288), Expect = 5e-26

Table 1E . Domain Analysis of NOV1

gnl|Pfam|pfam00094, vwd, von Willebrand factor type D domain.

CD-Length = 158 residues, 99.4% aligned

Score = 97.1 bits (240), Expect = 2e-20

Table 1F. Domain Analysis of NOV1

gnl|Smart|smart00231, FA58C, Coagulation factor 5/8 C-terminal domain, discoidin domain; Cell surface-attached carbohydrate-binding domain, present in eukaryotes and assumed to have horizontally transferred to eubacterial genomes.

CD-Length = 135 residues, 99.3% aligned

Score = 63.2 bits (152), Expect = 3e-10

Table 1G. Domain Analysis of NOV1

gnl|Smart|smart00209, TSP1, Thrombospondin type 1 repeats; Type 1 repeats in thrombospondin-1 bind and activate TGF-beta.

CD-Length = 51 residues, 100.0% aligned

Score = 53.1 bits (126), Expect = 3e-07

Table 1H. Domain Analysis of NOV1

gnl|Pfam|pfam00057, ldl_recept_a, Low-density lipoprotein receptor domain class A

CD-Length = 39 residues, 94.9% aligned

Score = 50.4 bits (119), Expect = 2e-06

The disclosed NOV1 protein contains a thrombospondin type I repeat domain which
 5 are found in the thrombospondin protein and is repeated 3 times. A number of proteins
 involved in the complement pathway (properdin, C6, C7, C8A, C8B, C9) as well as
 extracellular matrix protein like mindin, F-spondin, SCO-spondin and even the
 circumsporozoite surface protein 2 and TRAP proteins of Plasmodium contain one or more
 instance of this repeat. It has been involved in cell-cell interaction, inhibition of angiogenesis,
 10 apoptosis. The intron-exon organisation of the properdin gene confirms the hypothesis that the
 repeat might have evolved by a process involving exon shuffling. A study of properdin
 structure provides some information about the structure of the thrombospondin type I repeat.

The disclosed NOV1 nucleic acid of the invention encoding a DJ0751H13.1 PROTEIN
 -like protein includes the nucleic acid whose sequence is provided in Table 1A or a fragment
 15 thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may
 be changed from the corresponding base shown in Table 1A while still encoding a protein that
 maintains its DJ0751H13.1 PROTEIN like activities and physiological functions, or a
 fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences
 are complementary to those just described, including nucleic acid fragments that are
 20 complementary to any of the nucleic acids just described. The invention additionally includes
 nucleic acids or nucleic acid fragments, or complements thereto, whose structures include

chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 27 percent of the bases may be so changed.

The disclosed NOV1 protein of the invention includes the DJ0751H13.1 PROTEIN -like protein whose sequence is provided in Table 1B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 1B while still encoding a protein that maintains its DJ0751H13.1 PROTEIN -like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 0 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this DJ0751H13.1 PROTEIN -like protein (NOV1) may function as a member of a "DJ0751H13.1 PROTEIN family". Therefore, the NOV1 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV1 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in cancer including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the DJ0751H13.1 PROTEIN -like protein (NOV1) may be useful in gene therapy, and the DJ0751H13.1 PROTEIN -like protein (NOV1) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, transplantation, diabetes, autoimmune disease, renal artery stenosis, interstitial nephritis,

glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, hypercalcaemia, Lesch-Nyhan syndrome, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, cancers, and/or other pathologies and disorders. For example, a cDNA encoding the transmembrane receptor DJ0751H13.1 PROTEIN -like protein may be useful in transmembrane receptor DJ0751H13.1 PROTEIN therapy, and the transmembrane receptor DJ0751H13.1 PROTEIN -like protein may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, transplantation, diabetes, autoimmune disease, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, hypercalcaemia, Lesch-Nyhan syndrome, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, cancers, and other diseases, disorders and conditions of the like. Also since this gene is expressed at a measurably higher level in several cancer cell lines (including breast cancer, CNS cancer, colon cancer, gastric cancer, lung cancer, melanoma, ovarian cancer and pancreatic cancer), it may be useful in diagnosis and treatment of these cancers. The NOV1 nucleic acid encoding the DJ0751H13.1 PROTEIN -like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV1 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV1 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV1 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for

functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV2

A disclosed NOV2 nucleic acid of 893 nucleotides (also referred to as CG57558-01) encoding a Mac25/IGFBP7-like protein is shown in Table 2A. Putative untranslated regions upstream and/or downstream from the coding region, if any, are underlined, and the start and stop codons are in bold letters.

Table 2A. NOV2 nucleotide sequence (SEQ ID NO: 3).

GTACCTTAAAGACAACAAACAAGCAAACACAACCTTATAATTAAAAACATGCAAAGGGCT
CACCTTCCACTTCCTTCTGGTCCTGCTCCTCTTCTTCTCCTCTCCTGCCTCCTCTTCTC
CCTGTTCCATCAGACCTTCTGGGGCCCCCTTTCAATAAGCAGCTGCTGGCCGGCCAGCCCT
TGGGGGCAGGGCTGGAACCGGGGCAGGGGAGGCTGCGGGGCCACTCGCTGGAGAGGCAAA
CAGGAAGGACTGCCCCCTGAGCGCCAGGCTTCGGGGCCCGGAATCGCCGCCGCCGCCGCC
GCAGAGCTGCAGCTCGGGGCCGAGGGTAAGGAGGCGAGCCGGGAGCGGGAGGCCCGGGAG
AGCTCCGCGGGTCCCCGCGCCAGTCCCCAGCCGCGCCCGACCCCGCCGCCCGGGCCCT
AACGCGGCCCGCGAGGCCTACGCGGCGGCCGCCGTCACCGTGCTGGAGCCGCCGGCCTCC
GACCCCGAGCTGCAGCCCCGCGGAGCGCCCGCTGCCATCGCCGGGGTCCGGGGAGGGCGCC
CCGGTCTTCTCACGGGGCCTCGATCCCAGTGGGTGCTGCGGGGGGCGGAGGTGGTGTCTG
ACGTGCCCGGGCGGGGGCCCTCCCCGAGCCACACTGTACTGGGAGAAGGACGGGATGGCC
CTGGACGAAGTGTGGGACAGCAGCCACTTCGCGCTCCAGCCGGGGCCGCGCCGAGGACGGC
CCCGGCGCGAGCCTGGCACTGCGCATCTGGCGGCTCGGCTGCCGATTCCGGCGTCTAC
GTGTGCCACGCCGCAACGCGCACGGCCACGCGCAGGCGGGGGCGCTGCTCCAGGTGCTG
ACCCACACCTTCTGCGCCAAGACAGCCCT**TA**ACCAAGGCCAGAAAGGGTAG

In a search of public sequence databases, the NOV2 nucleic acid sequence, located on chromosome 2 has has 564 of 779 bases (72%) identical to a gb:GENBANK-ID:S56581|acc:S56581.1 mRNA from Rattus sp. (alpha inhibin gene {5' region} [rats, Genomic, 2141 nt]). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV2 polypeptide (SEQ ID NO:4) encoded by SEQ ID NO:3 has 274 amino acid residues and is presented in Table B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV2 has a signal peptide and is likely to be localized extracellularly with a certainty of 0.3700. The most likely cleavage site for a NOV2 peptide is between amino acids 32 and 33.

Table2B. Encoded NOV protein sequence (SEQ ID NO:4).

MQRAHLPLPSGPAPLLPPLLPLLPVPSDLLGPLSISSCWPASPWGQGWNRGRGGCGATR
WRGKQEGLPPEQASGPGIAAAAAELQLGAEGKEASREREARESSAGPRAQSPAAPRPR
RPGPNAAAGEAYAAAAVTVLEPPASDPELQPAERPLSPGSGEGAPVFLTGPRSQWVLRGA
EVLTCRAGGLPEPTLYWEKDGMADEVWDSHFALQPGRAEDGPGASLALRILARLPD
SGVYVCHARNAHGHQAQAGALLQVLTPTFLPPRQP

A search of sequence databases reveals that the NOV2 amino acid sequence has 80 of 266 amino acid residues (30%) identical to, and 112 of 266 amino acid residues (42%) similar to, the 277 amino acid residue ptmr:SPTREMBL-ACC:Q07822 protein from Homo sapiens (Human) (MAC25 PROTEIN). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV2 is expressed in at least brain, ovary, breast, testis. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV2 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 2C.

Table 2C. BLAST results for NOV2					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 14734071 ref XP_051017.1 (XM_051017)	KIAA0657 protein [Homo sapiens]	1025	143/150 (95%)	143/150 (95%)	1e-66
gi 13938170 gb AAH07201.1 AAH07201 (BC007201)	(protein for IMAGE:2961284) [Homo sapiens]	1044	143/150 (95%)	143/150 (95%)	4e-66
gi 18552587 ref XP_087161.1 (XM_087161)	(protein for IMAGE:2961284) [Homo sapiens]	180	50/58 (86%)	51/58 (87%)	4e-17
gi 9623317 gb AAF90112.1 AF254363.1 (AF254363)	stretchin-MLCK [Drosophila melanogaster]	281	35/105 (33%)	50/105 (47%)	7e-8
gi 17559576 ref NP_504582.1 (NM_072181)	titin [Caenorhabditis elegans]	2783	39/120 (32%)	57/120 (47%)	1e-7

Table 2D lists the domain descriptions from DOMAIN analysis results against NOV 2. This indicates that the NOV sequence has properties similar to those of other proteins known to contain this domain.

Table 2D. Domain Analysis of NOV2

gnl|Smart|smart00409, IG, Immunoglobulin
CD-Length = 86 residues, 100.0% aligned
Score = 57.4 bits (137), Expect = 1e-09

Mac25 is a follistatin (FS)-like protein that has a growth-suppressing effect on a p53-deficient osteosarcoma cell line (Saos-2). The protein exhibits a strong homology to FS, an activin-binding protein, and part of its sequence includes the consensus sequence of the member of the Kazal serine protease inhibitor family. The mac25 protein was localized in the cytoplasm and secreted into culture medium (1). Addition of recombinant mac25 protein (10-7 M) into the culture medium induced significant suppression of the growth of human cervical carcinoma cells (HeLa) and murine embryonic carcinoma cells (P19), as well as osteosarcoma cells (Saos-2). The mac25 protein was co-immunoprecipitated with activin A, a result that suggests that mac25 may be a secreted tumor-suppressor that binds activin A. The mac25 exhibits homology to insulin-like growth factor-binding proteins (IGF-BPs) and to fibroblast growth factor receptor. The multi-functional nature of mac25 protein may be important for growth-suppression and/or cellular senescence.

The disclosed NOV2 nucleic acid of the invention encoding a Mac25/IGFBP7-like protein includes the nucleic acid whose sequence is provided in Table 2A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 2A while still encoding a protein that maintains its Mac25/IGFBP7-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 2B percent of the bases may be so changed.

The disclosed NOV2 protein of the invention includes the Mac25/IGFBP7-like protein whose sequence is provided in Table 2B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table

2B while still encoding a protein that maintains its Mac25/IGFBP7-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 70 percent of the residues may be so changed.

5 The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this Mac25/IGFBP7-like protein (NOV2) may function as a member of a "Mac25/IGFBP7 family". Therefore, the NOV2 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

15 The NOV2 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the Mac25/IGFBP7-like protein (NOV2) may be useful in gene therapy, and the Mac25/IGFBP7-like protein (NOV2) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neurodegeneration, fertility, hypogonadism, endometriosis, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, graft versus host disease, or other pathologies or conditions. The NOV2 nucleic acid encoding the Mac25/IGFBP7-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

30 NOV2 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV2 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV2 proteins have multiple hydrophilic regions, each of

which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

5 NOV3

A disclosed NOV3 nucleic acid of 1703 nucleotides (also referred to as CG57560-01) encoding a Calmodulin Binding Protein Kinase-like protein is shown in Table 3A. Putative untranslated regions upstream and/or downstream from the coding region, if any, are underlined, and the start and stop codons are in bold letters.

10

Table 3A. NOV3 nucleotide sequence (SEQ ID NO:5).
<p>CCAGGTTGGGGTCTCCCAAAGCAGCCCCCTATGTTGTGGGGAAGTGGAGAGTAGTACAGC TAAGCCAGACCCCATTTGTGCCCCGAGGTTAGAGCCTGGCAATGCCGTTTGGGTGTGTGAC TCTGGGCGACAAGAAGAACTATAACCAGCCATCGGAGGTGACTGACAGATATGATTTGGG ACAGGTCATCAAGACGGAGGAGTTTTGTGAAATCTTCCGGGCCAAGGACAAGACGACAGG CAAGCTGCACACCTGCAAGAAGTTCCAGAAGCGGGACGGCCGCAAGGTGCGGAAAGCTGC CAAGAACGAGATAGGCATCCTCAAGATGGTGAAGCATCCCAACATCCTACAGCTGGTGGG TGTGTTTGTGACCCGCAAGGAGTACTTTATCTTCTGAGCTGGCCACGGGGAGGGAGGT GTTTGGTACTGGATCCTGGACCAGGGCTACTACTCGGAGCGAGACACAAGCAACGTGGTACG GCAAGTCTGGAGGCCGTTGGCTATTTGCACTCACTCAAGATCGTGCACAGGAATCTCAA GCTGGAGAACCTGGTTTACTACAACCGGCTGAAGAACTCGAAGATTGTATCATGAGTACTT CCATCTGGCTAAGCTAGAAAATGGCCTCATCAAGGAGCCCTGTGGGACCCCCGAGTATCT GGCCCCAGAGGTGGTAGGCCGGCAGCGGTATGGACGCCCTGTGGACTGCTGGGCCATTGG AGTCATCATGTACATCCTGCTTTCAGGCAACCCACCTTCTATGAGGAGGTGGAAGAAGA TGATTATGAGAACCATGATAAGAATCTTCCGCAAGATCCTGGCTGGTACTATGAGTT TGACTCTCCATATTGGGATGATATTTTCGAGGAGCCAAAGACCTGGTCACAAGGCTGAT GGAGGTGGAGCAAGACCAGCGGATCACTGCAGAAGAGGCCATCTCCCATGAGTGGATTTC TGGCAATGCTGCTTCTGATAAGAATCAAGGATGGTGTCTGTGCCAGATTGAAAAGAA CTTTGCCAGGGCCAAGTGAAGAAGGCTGTCCGAGTGACCAACCTCATGAAACGGCTCCG GGCACCAGAGCAGTCCAGCACGGCTGCAGCCAGTCCGGCCTCAGCCACAGACACTGCCAC CCCCGGGGCTGCAGGTGGGGCCACAGCTGCAGCTGCGAGTGGAGCTACCTCAGCCCCCTGA GGGTGATGCTGCTCGTGCTGCAAAGAGTGATAATGTGGCCCCCGCAGACCGTAGTGCCAC CCAGCCACAGATGGAAGTGCCACCCAGCCACTGATGGCAGTGTACCCAGCCACCCGA TGGAAGCATCACTCCAGCCACTGATGGGAGTGTACCCAGCCACTGACAGGAGCGCTAC TCCAGCCACTGATGGGAGAGCCACACCAGCCACAGAAGAGAGCACTGTGCCACCACCCCA AAGCAGTGCCATGCTGGCCACCAAGGCAGCTGCCACCCCTGAGCCGGCTATGGCCAGCC GGACAGCACAGCCCCAGAGGGGCCACAGGCCAGGCTCCACCCTCTAGTAAAGGGGAAGA GGCTGCTGGTTATGCCAGGAGTCTCAAAGGGAGGAGGCCAGCTGAGTAGGCAGCCTGGT GAGGGGGGGCAGGGGATGGGCAGGAGGGTGGGAGAGTGGATGAGGGGCTTCTCACTGTAC ATAGAGTCACTGGCATGATGCCC</p>

In a search of public sequence databases, the NOV3 nucleic acid sequence, located on chromosome 3 has 1015 of 1158 bases (87%) identical to a gb:GENBANK-
ID:RATCBVA|acc:L22557.1 mRNA from Rattus norvegicus (Rattus norvegicus vesicla-
associate calmodulin-binding protein mRNA, complete cds). Public nucleotide databases
include all GenBank databases and the GeneSeq patent database.

The disclosed NOV3 polypeptide (SEQ ID NO:6) encoded by SEQ ID NO:3 has 501 amino acid residues and is presented in Table 3B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV3 has no signal peptide and is likely to be localized in the in the cytoplasm with a certainty of 0.4500.

Table 3B. Encoded NOV3 protein sequence (SEQ ID NO:6).

MPFGCVTLGDKKNYNQPSSEVTDRLGQVIKTEEFCEIFRAKDKTGKLHTCKKFQKRDG
RKVRKAAKNEIGILKMVKHPNQLQVDVVFVTRKEYFIFLELATGREVFDWILDQGYYSER
DTSNVVRQVLEAVAYLHSLKIVHRNLKLENLVYVNRKNSKIVISDFHLAKLENGLIKEP
CGTPEYLAPVGRQRYGRPVDCWAIGVIMYILLSGNPPFYEEVEEDDYENHDKNLFKRI
LAGDYEFDSPYWDDISQAADLVTRLMEVEQDQRITAEAAISHWISGNAASDKNIKDG
CAQIEKNFARAKWKKAVRVTTLMKRLRAPEQSSTAAQASATDTATPGAAGGATAAAAS
GATSAPEGDAARAASDNVAPADRSATPATDGSATPATDGSVTPATDGSITPATDGSVTP
ATDRSATPATDGRATPATEESTVPTTQSSAMLATKAAATPEPAMAQPDSTAPEGATGQAP
PSSKGEEAAGYAQESQREEAS

A search of sequence databases reveals that the NOV3 amino acid sequence has 1015 of 1158 amino acid residues (87%) identical to, and 1015 of 1158 amino acid residues (87%) similar to, the 3655 amino acid residue gb:GENBANK-ID:RATCBVA|acc:L22557.1 protein from Rattus norvegicus (Rattus norvegicus vesicla-associate calmodulin-binding protein mRNA, complete cds). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV3 is expressed in at least Bone Marrow, Brain, Hypothalamus, Thalamus. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV3 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 3C.

Table 3C. BLAST results for NOV3

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 16924331 gb AAH17363.1 AAH17363 (BC017363)	protein MGC8407 [Homo sapiens]	501	501/501 (100%)	501/501 (100%)	0.0
gi 13129008 ref NP076951.1 (NM_024046)	protein MGC8407 [Homo sapiens]	501	500/501 (99%)	500/501 (99%)	0.0
gi 17160946 gb AAH17634.1 AAH17634 (BC017634)	Similar to vesicle- associated calmodulin- binding protein [Mus musculus]	512	468/512 (91%)	476/512 (92%)	0.0

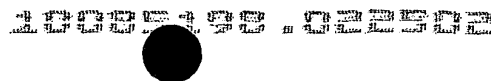
nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 13 percent of the bases may be so changed.

The disclosed NOV3 protein of the invention includes the Calmodulin Binding Protein Kinase-like protein whose sequence is provided in Table 3B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table B while still encoding a protein that maintains its Calmodulin Binding Protein Kinase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 13 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this Calmodulin Binding Protein Kinase-like protein (NOV3) may function as a member of a "Calmodulin Binding Protein Kinase family". Therefore, the NOV3 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV3 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the Calmodulin Binding Protein Kinase-like protein (NOV3) may be useful in gene therapy, and the Calmodulin Binding Protein Kinase-like protein (NOV3) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will



have efficacy for treatment of patients suffering from hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, graft versus host disease, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neurodegeneration, or other pathologies or conditions. The NOV3 nucleic acid encoding the Calmodulin Binding Protein Kinase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV3 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV3 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV3 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV4

NOV4 includes two TRANSIENT RECEPTOR POTENTIAL-RELATED PROTEIN-like proteins disclosed below. The disclosed sequences have been named NOV4a and NOV4b.

NOV4a

A disclosed NOV4a nucleic acid of 4877 nucleotides (also referred to as CG57547-01) encoding a TRANSIENT RECEPTOR POTENTIAL-RELATED PROTEIN-like protein is shown in Table 4A. Putative untranslated regions upstream and/or downstream from the coding region, if any, are underlined, and the start and stop codons are in bold letters.

Table 4A. NOV4 nucleotide sequence (SEQ ID NO:7).	
<u>AGTCCCCAGCCCCGTGCGCCGGCGGAGGCGGGCGCGGTCCTGTGGCCAGTCACCC</u>	
<u>GGAGGAGTTGGTCGCACAATTATGAAAGACTCGGCTTCTGCTGCTAGCGCCGGAGCTGAG</u>	
<u>TTAGTCCTGAGAAGGTTTCCCTGGGCGTTCCTTGTCGGCCCTCTGCTGCCGCCCTCCGGAG</u>	
<u>ACGCTTCCCGATAGATGGCTACAGGCCGCGGAGGAGGAGGTGGAGTTGCTGCCCTTC</u>	
<u>CGGAGTCCGCCCCGTGAGGAGAATGTCCCAAGAAATCCTGGATAGAAAGCACTTTGACCAA</u>	
<u>GAGGGAATGTGTATATATTATACCAAGTTCCAAGGACCCTCACAGGTGCCTTCCAGGATG</u>	
<u>TCAAATTTGTCAGCAACTCGTCAGGTGTTTTTGTGGTCGCTTGGTCAAGCAACATGCTTG</u>	
<u>TTTTACTGCAAGTCTTGCCATGAAATACTCAGATGTGAAATTGGGTGACCATTTAATCA</u>	

GGCAATAGAAGAATGGTCTGTGGAAAAGCATACAGAACAGAGCCCCAACGGATGCTTATGG
 AGTCATAAAATTTTCAAGGGGGTTCTCATTCTTACAGAGCTAAGTATGTGAGGCTATCATA
 TGACACCAAACCTGAAGTCATTCTGCAACTTCTGCTTAAAGAATGGCAAATGGAGTTACC
 CAAACTTGTATCTCTGTACATGGGGGCATGCAGAAATTTGAGCTTCACCCACGAATCAA
 GCAGTTGCTTGGAAAAGGTCTTATTAAAGCTGCAGTTACAAGTGGAGCCTGGAGTTTAAAC
 TGGAGGAGTAAACACAGGTGTGGCAAAACATGTTGGAGATGCCCTCAAAGAACATGCTTC
 CAGATCATCTCGAAAGATTGCACTATCGGAATAGCTCCATGGGGAGTGATTGAAAACAG
 AAATGATCTTGTGGGAGAGATGTAGTTGCTCCTTATCAAACCTTATTGAACCCCTGAG
 CAAATTGAATGTTTTGAATAATCTGCATTCCCATTTTCATATTGGTGGATGATGGCACTGT
 TGGAAAGTATGGGGCGGAAGTCAGACTGAGAAGAGAACTTGAAAAAATATTAATCAGCA
 AAGAATTATGCTATTGGCCAGGGTGTCCCTGTGGTGGCACTTATATTGAGGGTGGGCC
 AAATGTTTATCCTCACAGTTCTTGAATACCTTCAGGAAAGCCCCCTGTTCCAGTAGTTGT
 GTGTGAAGGAACAGGCAGAGCTGCAGATCTGCTAGCGTATATTCTATAAACAAACAGAAGA
 AGGAGGGAATCTTCTGATGCAGCAGAGCCCGATATTATTTCCACTATCAAAAAACATT
 TAACTTTGGCCAGAATGAAGCACTTCATTTATTTCAAACACTGATGGAGTGCATGAAAAG
 AAAGGAGCTTATCACTGTTTTCCATATTGGGTGAGATGAACATCAAGATATAGATGTAGC
 AATACTTACTGCAC'TGCTAAAAGGTACTAATGCATCTGCATTGACCAGCTTATCCTTAC
 ATTGGCATGGGATAGAGTTGACATTGCCAAAAATCATGTATTGTTTTATGGACAGCAGTG
 GCTGGTAGGATCCTTGGAAACAAGCTATGCTTGATGCTCTTGTAATGGATAGAGTTGCATT
 TGTAAAACCTTCTTATTGAAAATGGAGTAAGCATGCATAAATTCTTACCATTCCGAGACT
 GGAAGAACTTTACAACACTAAACAAGGTCCAATAATCCAATGCTGTTTCATCTTGTTCG
 AGACGTCAAACAGGGAAATCTTCTCCAGGATATAAGATCACTCTGATTGATATAGGACT
 GTTATTGAATATCTCATGGGAGGAACCTACAGATGCACCTATACTAGGAAACGTTTTTCG
 ATTAATATATAATAGTCTTGGTGGAAATAATCGGAGGTCTGGCCGAAATACCTCCAGCAG
 CACTCCTCAGTTGCCAAAGAGTCATGAATCTTTTGGCAATAGGGCAGATAAAAAGGAAAA
 AATGAGGCATAACCATTTCATTAAGACAGCACAGCCCTACCGACCAAAGGTAGATACAGT
 TATGGAAGAAGGAAAGAAAGAAAAGAACCAAGATGAAATTGTAGACATTGATGATCCAGA
 AACCAAGCGCTTCTCTTATCCACTTAATGAACCTTTAATTTGGGCTTGCCTTATGAAGAG
 GCAGGTGATGGCCCGTTTTTTATGGCAACATGGTGAAGAATCAATGGCTAAAGCATTAGT
 TGCTGTGAAGATCTATCGTTCAATGGCATATGAAGCAAAGCAGAGTGACCTGGTAGATGA
 TACTTCAGAAGAACTAAAACAGTATTCCAGTGATTTTGGTCACTTGGCCGTTGAATTATT
 AGAACAGTCTTTCAGACAAGATGAAACCATGGCTATGAAATTGCTCACTTATGAAGTGA
 GAAGTGGAGTAATTCACCTGCCTTAAGTTAGCAGTTTCTTCAAGACTTAGACCTTTTGT
 AGCTGAAGACTGTACACAATGTTGTTATCTGATATGTGGATGGGAAGGCTGATGATCCAG
 GAAAAATTCCTGGTACAAGGTAATACTAAGCATTTTAGTTCCACCTGCCATATTGCTGTT
 AGAGTATAAAACTAAGGCTGAAATGTCCCATATCCACAATCTCAAGATGCTCATCAGAT
 GACAATGGATGACAGCGAAAACAACCTTTCAGAACATAACAGAAGAGATCCCCATGGAAGT
 GTTTAAAGAAGTACGGATTTTGGATAGTAATGAAGGAAAGAATGAGATGGAGATACAAAT
 GAAATCAAAAAGCTTCCAATTACGCGAAAGTTTATGCCTTTTATCATGCACCAATTGT
 AAAATTCTGGTTTAAACAGTTGGCATATTTAGGATTTCTGATGCTTTATACATTTGTGGT
 TCTGTACAAATGGAACAGTTACCTTTCAGTTCAAGAATGGATTGTTATTGCTTATATTTT
 TACTTATGCCATTGAGAAAGTCCGTGAGGTATTTATGTCTGAAGCTGGGAAAGTAAACCA
 GAAGATTAAAGTATGGTTTAGTGATTACTTCAACATCAGTGATACAATTGCCATAATTTT
 TTTCTTCATTGGATTTGGACTAAGATTTGGAGCAAATGGAACCTTGCATGATGATGA
 TAATCATGTTTTTGTGGCTGGAAGATTAATTTACTGTCTTAACATAATATTTTGGTATGT
 GCGTTTGTAGATTTTCTAGCTGTAAATCAACAGGCAGGACCTTATGTAATGATGATTGG
 AAAAATGGTGGCCAATATGTTCTACATTGTAGTGATTATGGCTCTTGTATTACTTAGTTT
 TGGTGTTCAGAAAGGCAATACTTTATCCTCATGAAGCACCATCTTGGACTCTTGCTAA
 AGATATAGTTTTTCAACCCATACTGGATGATTTTGGTGAAGTTTATGCATACGAAATTGA
 TGTGTGTGCAAATGATTCTGTATCCCTCAAATCTGTGGTCTGGGACGTGGTTGACTCC
 ATTTCTTCAAGCAGTCTACCTCTTTGTACAGTATATCATTATGGTTAATCTTCTTATTGC
 ATTTTTTCAGCAATGTGTATTTACAAGTGAAGGCAATTTCCAATATTGTATGGAAGTACCA
 GCGTTATCATTTTATTATGGCTTATCATGAGAAACCAGTTCTGCCTCCTCCACTTATCAT
 TCTTAGCCATATAGTTTCTCTGTTTTGCTGCATATGTAAGAGAAGAAAGAAAGATAAGAC
 TTCCGATGGACAGAACTTTTCTTAACAGAAGAAGATCAAAGAAACTTCATGATTTTGA
 AGAGCAGTGTGTTGAAATGTATTTCAATGAAAAAGATGACAAATTTTCACTCTGGGAGTGA
 AGAGAGAATTCGTGTCACTTTTGAAGAGTGGAAACAGATGTGCATTGAGATTAAAGAAGT
 TGGAGATCGTGTCACTACATAAAAAGATCATTACAATCATTAGATTCTCAAATTGGCCA
 TTTGCAAGATCTTTCAGCCCTGACGGTAGATACATTAAAAACACTCACTGCCCAGAAAGC
 GTCGGAAGCTAGCAAAGTTTCAATGAATCACACGAGAAGTGAAGCATTTCCAAACACTT
 GGCTCAAAACCTTATTGATGATGGTCTGTAAAGACCTTCTGTATGGAAAAAGCATGGTGT
 TGTAAATACACTTAGCTCCTCTCTTCTCAAGGTGATCTTGAAAGTAATAATCCTTTTCA

TTGTAATATTTTAAATGAAAGATGACAAAGATCCCCAGTGTAATATATTTGGTCAAGACTT
ACCTGCAGTACCCAGAGAAAAGAATTTAATTTTCCAGAGGCTGGTTCCTCTCTGGTGC
CTTATTTCCCAAGTGCTGTTTTCCCTCCAGAACTGCGACAGAGACTACATGGGGTAGAACT
CTTAAAAATATTTAATAAAAAATCAAAAATTAGGCAGTTCATCTACTAGCATACCACATCT
GTCATCCCCACCAACCAAAATTTTTGTTAGTACACCATCTCAGCCAAGTTGCAAAAGCCA
CTTGGAACCTGGAACCAAGATCAAGAACTGTTTGCTCTAAAGCTACAGAAGGAGATAA
TACAGAATTTGGAGCATTTGTAGGTCACAGAGATAGCATGGATTTACAGAGGTTTAAAGA
AACATCAAACAAGATAAAATTGCAGAATAACAATACTTCTGAAAACACTTTGAAACGAGT
GAGTTCTCTGCTGGATTTACTGACTGTCACAGAACTTCCATTCTGTTCATTCAAACA
AGCAGAAAAATCAGTAGAAGGCCATCTACCGAAGACACTCATGAAGTAGATTCCAAAGC
AGCTTTATTACTGAAGGATTGGTTACAAGATAGACCATCAAACAGAGAAATGGGTCTCAC
TTCTCCATTTAAGCCAGCTATGGATACAAATTACTATTATTAGCTGTGGAAAGAAATAA
CTTGATGAGGTTATCAGAGAGCATTCATTTACACCTGTGCCTCCAAGAGGGAGCCTGT
CACAGTGTATCGTTTGGAAGAGAGTTACCCCAACATACTAAATAACAGCATGTTCTCTTG
GTCACAACCTAGGCCTCTGTGCCAAAATAGAGTTTTTAAGCAAAGAGGAGATGGGAGGAGG
TTTACGAAGAGCTGTCAAAGTACAGTGTACCTGGTCAGAACATGATATCCTCAAATCAGG
GCATCTTTATATTATCAAATCTTTTCTCCAGAGGTGGTTAATACATGGTCAAGTATTTA
CAAAGAAGATACAGTTCTGCATCTCTGTCTGAGAGAAATTCAACAACAGAGAGCAGCACA
AAAGCTTACGTTTGCCTTTAATCAAATGAAACCCAAATCCATACCATATTCTCCAAGGTT
CCTTGAAGTTTTCTGCTGTATTGCCATTACAGCAGGACAGTGGTTTGTGTGGAAGAATG
TATGATGGAGAATTTAGAAAATACAACAATAATAATGGAGATGAGATTATCCAACTAA
TACTCTGGAAGAGATCATGCTAGCCTTTAGCCACTGGACTTACGAATATACAAGAGGGGA
GTTACTGGTACTTGATTTGCAAGGTGTTGGTGAAAATTTGACTGACCCATCTGTGATAAA
AGCAGAAGAAAAGAGATCCTGTGATATGGTTTTTGGCCAGCAAATCTAGGAGAAGATGC
AATTAATAAATCTCAGAGCAAACATCACTGTAATCTTGCTGTAGAAAGCTTAAACTTCC
AGATCTGAAGAGGAATGATTATACGCCTGATAAAATTATATTTCTCAGGATGAGCCTTC
AGATTTGAATCTTCAGCCTGGAAATCCACCAAAGAATCAGAATCAACTAATTCTGTTTCG
TCTGATGTTATTAATTAATATTACTGAATCATTTGTTTTGCCTGCACCTCACAGAA

In a search of public sequence databases, the NOV4a nucleic acid sequence, located on chromosome 15 has 4374 of 4825 bases (90%) identical to a gb:GENBANK-ID:AF149013|acc:AF149013.1 mRNA from Mus musculus (Mus musculus transient receptor potential-related protein (ChaK) mRNA, complete cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV4a polypeptide (SEQ ID NO:8) encoded by SEQ ID NO:7 has 1856 amino acid residues and is presented in Table 4B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV4a has no signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.6000.

Table 4B. Encoded NOV4a protein sequence (SEQ ID NO:8).	
MSQKSWIESTLTKRECVYIIPSSKDPHRCPLPGCQICQQLVRCFCGRLVKQHACFTASLAM	
KYSDVKLGDHFNQAIEEWSVEKHTEQSPTDAYGVINFQGGSHSYRAKYVRLSYDTKPEVI	
LQLLLKEWQMELPKLVISVHGGMQKFELHPRIKQLLGKGLIKAAVTGAWILTGAVNTGV	
AKHVGDAKKEHASRSSRKICTIGIAPWGVNIENRNDLVGRDVPVAPYQTLNPLSKLNVLN	
LHSHFILVDDGTGKYGAEVRLRRELEKTINQQRHAIGQGVPPVVALIFEGGPNVILTVL	
EYLQESPPVPVVVCEGTGRAADLLAYIHKQTEEGGNLPDAAEPDIISTIKKTFNFGQNEA	
LHLFQTLMECMKRKELITVFHIGSDEHQDIDVAILTALLKGTNASAFDQLILTAWDRVD	
IAKNHVFVYQQWLVGSLQAMLDALVMDRVAFVKLLIENGVS MHKFLTIPRLEELYNTK	
QGPTNPMLFHLVRDVKQGNLPPGYKITLIDIGLVIEYLMGGTYRCTYTRKFRFLIYNSLG	
GNNRRSGRNTSSSTPQLRKSHESFGNRADKKEKMRHNHFIKTAQPYRPKVDVMEEGKKK	
RTKDEIVDIDDPETKRFPYPLNELLIWACL MKRQVMARFLWQHGEESMAKALVACKIYRS	
MAYEAKQSDLVDDTSEELKQYSSDFGQLAVELLEQSFRQDETAMKLLTYELKNWSN	
STCLKLA VSSRLRPVHAHTCTQMLLSDMWMGR LNMKNSWYK VILSILVPPAILLLEYKTKAE	

MSHIPQSDAHQMTMDDSENNFQNI TEEI P MEVFKEVRILDSNEGKNEMEI QMKS KKLPI
TRKFYAFYHAPIVKFWFNTLAYLGFLMLYTFVVLVQMEQLPSVQEWIVIAIYIFTYAI EKV
REVMSEAGKVNQKIKVWFSDYFNISDTIAIISFFIGFLRFGAKWNFANAYDNHVFVAG
RLIYCLNII FWYVRLD FLAVNQAGPYVMMIGKVMANMFYIVVIMALVLLSFGVPRKAI
LYPHEAPSWTLAKDIVFHPYWMIFGEVYAYEIDVCSANDSVIPQICGPGTWLTPFLQAVYL
FVQYIIMVNLLIAFFSNVYLQVKAISNIVWKYQRYHFIMAYHEKPVLPPPLIILSHIVSL
FCCICKRRKKDKTSDGPELFLTEEDQKKLHDFEEQCVEMYFNEKDDKFHSGSEERIRVTF
ERVEQMCIQIKEVGDRVNYIKRSLQSLDSQIGHLQDLSALTVDTLKTLTAQKASEASKVH
NEITRELSISKHLAQNLI DDGPVRPSVWKKHGVVNTLSSSLPQGDLESNNPFHCNII LMKD
DKDPQCNIFGQDLPAVPQRKEFNFP EAGSSSGALFPSAVSPPELRQRLHGV ELLKIFNKN
QKLGSSSTSIPHLSSPPTKFFVSTPSQPSCKSHLETGTDQETVCSKATEGDNTEFGAFV
GHRDSMDLQRFKETS NKIKLQNNNTSENTLKRVS LAGFTDCHRTSIPVH SKQAEKISR
PSTEDTHEVDSKAALLLKDWLQDRPSNREMGLTSPFKPAMDTNYYS AVERNLMRLS QS
IPFTFPVPPRGPVTVYRLEESSPNILNNSMSSWSQLGLCAKIEFLSKEEMGGGLRRRAVKV
QCTWSEHDILKSGHLYI IKSFLPEVVNTWSS IYKEDTVLHLCLREIQQR AAQKLTFAFN
QMKPKSIPYSPRFLEVFLLYCHSAGQWFAVEECMTGEFRKYNNNGDEI IPTNTLEEIML
AFSHWTYEYTRGELLVLDLQGVGENLTDPSVIKAE EKRS CDMVFGPANLGEDA IKNFR AK
HHCNSCCRKLKLPDLKRNDYTPDKIIFPQDEPSDLNLQPGNSTKESESTNSVRLML

A search of sequence databases reveals that the NOV4a amino acid sequence has 1747
of 1863 amino acid residues (93%) identical to, and 1803 of 1863 amino acid residues (96%)
similar to the 1863 amino acid residue ptnr:SP TREMBL-ACC:Q9JLQ1 protein from Mus
5 musculus (Mouse) (TRANSIENT RECEPTOR POTENTIAL-RELATED PROTEIN). Public
amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV4a is expressed in at least Adrenal Gland/Suprarenal gland, Bone Marrow, Brain,
Bronchus, Cartilage, Colon, Hippocampus, Kidney, Liver, Lymph node, Skeletal Muscle,
Stomach, Substantia Nigra, Tonsils and Whole Organism. This information was derived by
10 determining the tissue sources of the sequences that were included in the invention including
but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE
sources.

NOV4b

A disclosed NOV4b nucleic acid of 5626 nucleotides (also referred to as CG57547-
15 02) encoding a TRANSIENT RECEPTOR POTENTIAL-RELATED PROTEIN-like protein is shown in
Table 4C. Putative untranslated regions upstream and/or downstream from the coding
region, if any, are underlined, and the start and stop codons are in bold letters.

Table 4C. NOV4b nucleotide sequence (SEQ ID NO:9).	
<p><u>AGTCCCCAGCCCCGTCGCCGGCGGAGGCGGGCGGGCGCGTTCCTGTGGCCAGTCA</u><u>CCCGGAGGAGTTGGTCGCACAATTATGAAAGACTCGGCTTCTGCTGCTAGCGCCGGAGCTGAG</u> <u>TTAGTCCTGAGAAGGTTTCCCTGGGCGTTCCTTGTCCGGCCTCTGCTGCCGCCTCCGGAG</u> <u>ACGCTTCCC</u>GATAGATGGCTACAGGCCGCGGAGGAGGAGGAGGTGGAGTTGCTGCCCTTC CGGAGTCCGCCCCGTGAGGAGAATGTCCAGAAATCCTGGATAGAAAGCACTTTGACCAA GAGGGAATGTGTATATATTATACCAAGTTC AAGGACCCTCAGGTCCTTCCAGGATG TCAAAATTTGTCAGCAACTCGTCAGGGTTTTTGTGGTCGCTTGGTCAAGCAACATGCTTGT</p>	

TTTACTGCAAGTCTTGCCATGAAATACTCAGATGTGAAATTGGGTGACCATTTTAATCAG
GCAATAGAAGAATGGTCTGTGAAAAAGCATACAGAACAGAGCCCCAACGGATGCTTATGGA
GTCATAAATTTTCAAGGGGGTCTCATTCTACAGAGCTAAGTATGTGAGGCTATCATAT
GACACCAAACTCTGAAGTCATTCTGCAACTCTGCCTTAAAGATGGCAATGGAGTATACCC
AAACTTGGTATCTGTACATGGGGCATGCGAAATTTGAGCTTCACCCACGAATCAAG
CAGTTGCTTGGAAAAGGTCTTTATTAAGCTGCAGTTACAACCTGGAGCCTGGATTTTAAC
GGAGGAGTAAACACAGGTACAGGTGTGGCAAAACATGTTGGAGATGCCCTCAAAGAACAT
GCTTCCAGATCATCTCGAAAGATTTGCACTATCGGAATAGCTCCATGGGGAGTGATTGAA
AACAGAAATGATCTTGTGGGAGAGATGTAAGAATTATTTATCAAACCTTATTGAACCCC
CTGAGCAAAATTGAATGTTTTGAATAATCTGCATTCCCATTTCATATTGGTGGATGAGC
ACTGTTGGAAAGTATGGGGCGGAAGTCAGACTGAGAAGAGAATGAAAAACATTAAT
CAGCAAAAGAATTCATATTGGCCAGGGTGTCCCTGTGGTGGCACTTATATTGAGGGTGGG
CCAAATGTTATCCTCACAGTTCTTGAATACCTTCAGGAAAGCCCCCTGTTCCAGTAGTT
GTGTGTGAAGGAACAGGCAGAGCTGCAGATCTGCTAGCGTATATTCTATAACAAACAGAA
GAAGGAGGGGATGCAGCAGAGCCCGATATTATTTCCACTATCAAAAAACATTTAACTTT
GGCCAGAATGAAGCACTTCATTTATTTCAAACACTGATGGAGTGCATGAAAAGAAAGGAG
CTTGTAACCTGTTTTCCATATTGGGTGCAGATGAACATCAAGATATAGATGTAGCAATACT
ACTGCACTGTCTAAAAGGTACTAATGCATCTGCATTGACAGCTTATCTTTACATTGGCA
TGGGATAGAGTTGACATTGCCAAAAATCATGTATTTGTTTATGGACAGCAGTGGCCATTG
CACTCCAGCCTGGGCAACAGAGTGAGACTCTCTCTCAAAAAAAAAAAAAACAAAAACAAAA
CAAAAAACAAAAACAAAAACCAACACCTAGAAATTGAGAGTTAGTTGGATCCTTGGAAACAA
GCTATGCTTGATGCTCTTGTAATGGATAGAGTTGCATTTGTAAAACCTTCTTATTGAAAT
GGAGTAAGCATGCATAAATCTTACCATTCCGAGACTGGAAGAACTTTACAACACTAAT
CTTCTCCAGGATATAAGATCACTCTGATTGATATAGGACTTGTATTGAATATCTCATG
GGAGGAACCTACAGATGCACCTATACTAGGAAACGTTTTTCGATTAATATATAATAGTCTT
GGTGGAAATAATCGGTTTTCTTCCAGGAGCCCAACCACTCGCACGGTAAATATTAGA
GACAAATCTCCTCATGCTTCTGGCAAGAAGAAGGGAAAAGAAGAAAAGAACCAAGATGAA
ATTGTAGACATTGATGATCCAGAAACCAAGCGCTTTCCTTATCCACTTAATGAACCTTTA
ATTTGGGCTTGCCCTTATGAAGAGGCAGGTCATGGCCCGTTTTTTATGGCAACATGGTGAA
GAATCAATGGCTAAAGCATTTGCTTGCCGTGAAGATCATCGTTCAATGGCATATGAAGCA
AAGCAGAGTGACCTGGTAGATGATACTTCAGAAGAACTAAAACAGTATTTCCAAGGATTTT
GGTCAGTTGGCCGTTGAATTATTAGAACAGTCCCTTCAGACAAGATGAAACCATGGCTATG
AAATTGCTCACTTATGAACTGAAGAAGTGGAGTAATTCAACCTGCCTTAAGTTAGCAGTT
TCTTCAAGACTTAGACCTTTTGTAGCTCACACCTGTACACAAATGTTGTTATCTGATATG
TGGATGGGAAGGCTGAATATGAGGAAAAATTCCTGGTACAAGGTAAATAAGCATTTTA
GTTCCACTTGCCATATTGCTGTGTAGAGTATAAAACTAAGGCTGAAATGTTCCCATATCCCA
CAATCTCAAGATGCTCATCAGATGACAATGGATGACAGCGAAAACAACAGTAATGAAGGA
AAGAATGAGATGGAGATACAAATGAAATCAAAAAAGCTTCCAATTACGCGAAAGTTTTAT
GCCTTTTATCATGCACCAATTGTAAAATTCCTGGTTTAAACAGTTGGCATATTTAGGATTT
CTGATGCTTTATACATTTGTGGTTCTTGTACAAATGGAACAGTTACCTTCAGTTCAAGAA
TGGATTGTATTAGCTTATATTTTATGCTATTGCCATTGAGAAAGTCCGTGAGATCTTTATG
TCTGAAGCTGGGAAAGTAAACCAGAAGATTAAAGTATGGTTTTAGTATTACTTCAACATC
AGTGATACAATTGCCATAAATTTCTTTCTTCATTGGATTTGGACTAAGATTTGGAGCAAAA
TGGAACCTTTGCAAATGCATATGATAATCATGTTTTTGTGGCTGGAAGATTAATTTACTGT
CTTAACATAATATTTTTGGTATGTGCGTTTGCTAGATTTTCTAGCTGTTAATCAACAGGCA
GGACCTTATGTAATGATGATTGGAAAAATGGTAAATATGTTCTACATTGTAGTGATTATG
GCTCTTGTATTACTTAGTTTTGGTGTTCAGAAAGGCAATACCTTATCCTCATGAAGCA
CCATCTTGGACTCTTGGCTAAGAGTATAGTTTTTCACCCATACTGGATGATTTTGGTGAA
GTTTTATGCATACGAAATTGATTGTGGTCTCGGGACGTGGTTGACTCCATTTCTTCAAGCA
GTCTACCTCTTTGTACAGTATATCATTATGGTTAATCTTCTTATTGCATTTTCAAGAGC
AATGTGTATTTACAAGTGAAGGCAATTTCCAATATTGTATGGAAGTACCAGCGTTATCAT
TTTATTATGGCTTATCATGAGAAACAGTTCTGCCTCCTCCACTTATCATTCTTAGCCAT
ATAGTTTCTCTGTTTTGCTGCATATGTAAGAGAAGAAAGATAAGACTTCCGATGGGA
CCAAGTAAAGATAGAATTTTTCTTAAACAGAAGAAGATCAAAAGAACTTCATGATTTTGAA
GAGCAGTGTGTTGAAATGTATTTCAATGAAAAGATGACAAATTTTCAATTCTGGGAGTGAA
GAGAGAATTCGTGTCACTTTTGAAAGAGTGGAAACAGAAGCCCATTCAGATTAAGAAGTT
GGAGATCGTGTCAACTACATAAAAAGATCATTACAATCATTAGATTCTCAAATTGGCCAT
TTGCAAGATCTTTCAGCCCTGACGGTAGATACATTAAAAACACTCACTGCCAGAAAGCG
TCGGAAGCTAGCAAAAGTTCAATGAATAACACACGAGAAGTGAACATTTCCAACACTGTG
GCTCAAAACCTTATGATGATGTTCTGTAAGACCTTCTGTATGGAAGAAAGCATGGTGT
GTAAATACACTCTAGCTCTCTCTTCTCAAGGTGATCTTGAAAGTAATAATCTTTTCAT
TGTAATATTTTAATGAAAGATGACAAAGATCCCCAGTGTAATATATTTGGTCAAGACTTA

CCTGCAGTACCCAGAGAAAAGAATTTAATTTTCCAGAGGCTGGTTCCTCTCTGCTGCC
TTATTCCCAAGTGCTGTTTCCCTCCAGAACTGCGACAGAGACTACATGGGGTAGAACTC
TTAAAAATATTTAATAAAAAATCAAAAATTAGGCAGTTCATCTACTAGCATACCACATCTG
TCATCCCCACCAACCAAAATTTTGTGTAGTACACCATCTCAGCCAAGTTGCAAAAGCCAC
TTGGAAGTGGAAACCAAGATCAAGAACTGTTTGTCTTAAAGCTACAGAAGGAGATAAT
ACAGAATTTGGAGCATTGTAGGTACACAGAGATAGCATGGATTTACAGAGGTTTAAAGAA
ACATCAAAACAAGATAAAATTCAGAAATAACAATACTTCTGAAAACACTTTGAAACGAGTG
AGTTCTCTGCTGGATTTACTGACTGTACAGAACTTCCATTCTGTTTATTCAAACAA
GCAGAAAAATCAGTAGAAGGCCATCTACCGAAGACACTCATGAAGTAGATTCCAAAGCA
GCTTTAATACCGGATTGGTTACAAGATAGACCATCAAACAGAGAAATGGGTCTCACTTCT
CCATTTAAGCCAGCTATGGATACAAATTACTATTATTAGCTGTGGAAGAAATAACTTG
ATGAGGTATACAGAGCATTCCATTACACCTGTGCCTCCAAGAGGGGAGCCTGTGACA
GTGTATCGTTTGAAGAGAGTTACCCCAACATACTAAATAACAGCATGTCTCTTGGTCA
CAACTAGGCCTCTGTGCCAAAATAGAGTTTTTAAGCAAAGAGGAGATGGGAGGAGTTTA
CGAAGAGCTGTCAAAGTACAGTGTACCTGGTCAGAACATGATATCCTCAAATCAGGGCAT
CTTTATATTATCAAATCTTTTCTCCAGAGGTGGTTAATACATGGTCAAGTATTTACAAA
GAAGATACAGTTCTGCATCTCTGTCTGAGAGAAATTAACAACAGAGAGCAGCACAAAAG
CTTACGTTTGCCTTTAATCAAATGAAACCCAAATCCATACCATATTCTCCAGGGGAGTTA
CTGGTACTTGATTGCAAGGTGTGGTGAAAATTTGACTGACCCATCTGTGATAAAAGCA
GAAGAAAAGAGATCCTGTGATATGGTTTTTGGCCAGCAAATCTAGGAGAAGATGCAATT
AAAAACTTCAGAGCAAAACATCACTGTAAATCTTGTCTGTAGAAAGCTTAACTTCCAGAT
CTGAAGAGGAATGATTATACGCCTGATAAAATTATATTTCTCAGGATGAGCCTTCAGAT
TTGAATCTTCAGCCTGGAAATTCACCAAAGAATCAGAATCAACTAATTCTGTTCTGTG
ATGTTATAATATTAATATTACTGAATCATTTGGTTTTGCCTGCACCTCACAGAAATGTTAC
TGTGTCACTTTTCCCTCGGGAGGAAATTGTTTGGTAAATAGAAAG

In a search of public sequence databases, the NOV4b nucleic acid sequence, located on chromosome 15 has 1134 of 1246 bases (91%) identical to a gb:GENBANK-
ID:AF149013|acc:AF149013.1 mRNA from Mus musculus (Mus musculus transient receptor
5 potential-related protein (ChaK) mRNA, complete cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV4b polypeptide (SEQ ID NO:10) encoded by SEQ ID NO:9 has 1815 amino acid residues and is presented in Table 4D using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV4b has no signal peptide and is
10 likely to be localized at the plasma membrane with a certainty of 0.6000.

Table 4D. Encoded NOV4b protein sequence (SEQ ID NO:10).	
MKDSASAASAGAEVLRRFPWAFVLRPLPPPETLPDRWLQAAEEEEVELLPFRSPPREE	
NVPEILDRKHFDQEGMCIYYTKFQGPSQVPSRMSNLSATRQGFGRVLVKQHACFTASLAM	
KYSDVKLGDHFNQAI E EWSVEKHTEQSPTDAYGVINFQGGSHSYRAKYVRLSYDTKPEVI	
LQLLLKEWQMELPKLVISVHGMQKFELHPRIKQLLGKGLIKA AVTTGAWILTGGVNTGT	
GVAKHVGDALKEHASRSSRKICTIGIAPWGVIENRNDLVGRDVR I IYQTLNPLSKLNLV	
NNLHSHF ILVDDGTGKYGA EVRLRRELEKTINQQR IHIHQGVPPVALIFEGGPNVILT	
LEYLQESPPVPVVVCEGTGRAADLLAYIHKQTEEGDAAE PD I I S I K K T F N F Q N E A L H	
LFQTLMECMKRKELVTVFHIGSDEHQDIDVA I L T A L L K G T N A S A F D Q L I L T L A W D R V D I A	
KNHV F V Y G Q Q W P L H S S L G N R V R L S L K K K K Q K Q K Q K Q K P T P R N S E L V G S L E Q A M L D A L V	
M D R V A F V K L L I E N G V S M H K F L T I P R L E E L Y N T N L P P G Y K I T L I D I G L V I E Y L M G G T Y R C T	
Y T R K R F R L I Y N S L G G N N R F S F Q E P N H T R T V N I R D K S P H A S G K K K G K K R T K D E I V D I D D P	
E T K R F P Y P L N E L L I W A C L M K R Q V M A R F L W Q H G E E S M A K A L V A C K I Y R S M A Y E A K Q S D L V D	
D T S E E L K Q Y S K D F G Q L A V E L L E Q S F R Q D E T M A M K L L T Y E L K N W S N S T C L K L A V S S R L R P F	
V A H T C T Q M L L S D M W M G R L N M R K N S W Y K V I L S I L V P P A I L L L E Y K T K A E M S H I P Q S Q D A H Q	
M T M D D S E N N S N E G K N E M E I Q M K S K K L P I T R K F Y A F Y H A P I V K F W F N T L A Y L G F L M L Y T F V	
V L V Q M E Q L P S V Q E W I V I A Y I F T Y A I E K V R E I F M S E A G K V N Q K I K V W F S D Y F N I S D T I A I I	

SFFIGFGLRFGAKWNFANAYDNHVFVAGRLIYCLNII FWYVRLLD FLAVNQAGPYVMMI
GKMVMNFYIVVIMALVLLSFGVPRKAILYPHEAPSWTLAKDIVFHPYWMIFGEVYAYEID
CGPGTWLTPFLQAVYLFVQYIIMVNLLIAFFKSNVYLQVKAI SNIVWKYQRYHFIMAYHE
KPVLPPLIILSHIVSLFCCICKRRKKDKTSDGSPKIELFLTEEDQKKLHDFEEQCVEMY
FNEKDDKFHSGSEERIRVTFERVEQKPIQIKEVGDRVNYIKRSLQSLDSQIGHLQDLSAL
TVDTLKTTLTAQKASEASKVHNEITRELSISKHLAQNLIDDGPVRPSVWKKHGVNTLS
LPQGDLESNNPFHCNILMKDDKDPQCNI FGQDLPAVPQRKEFNFP EAGSSSGALFPSAVS
PPELRQRLHGVELLKIFNKNQKLGSSSTSIPHLSSPPTKFFVSTPSQPSCKSHLETGTD
QETVCSKATEGDNTEFGAFVGHRSMDLQRFKETS NKIKLQNNNTSENTLKRVS
LAGFTDCHRTSIPVHSKQAEKISRRPSTEDTHEVDSKAALIPDWLQDRPSNREMG
LTSFPFKPAMDTNYYSAVERNNLMRLSQSIPFTPVP PRGEPVTVYRLEESSPNILNNS
MSSWSQLGLCAKIEFLSKEEMGGGLRRRAVKVQCTWSEHDILKSGHLYI IKSFLPEV
VNTWSSIYKEDTVLHLCLREIQQRAAQKLTFAFNQMKPKSIPYSPGELLVLDLQGV
GENLTDPSVIKAEKRSCDMVFGPANLGEDAIKNFRAKHHCNSCCRKLKLPDLKRNDY
TPDKIIFPQDEPSDLNLQPGNSTKSESTNSVRLML

A search of sequence databases reveals that the NOV4b amino acid sequence has 776 of 892 amino acid residues (86%) identical to, and 819 of 892 amino acid residues (91%) similar to, the 1863 amino acid residue ptnr:SPTREMBL-ACC:Q9JLQ1 protein from Mus musculus (Mouse) (TRANSIENT RECEPTOR POTENTIAL-RELATED PROTEIN). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV4b is expressed in at least adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea and uterus. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV4a polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 4E.

Table 4E. BLAST results for NOV4					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 13959785 gb AAK44211.1 (AY032950)	LTRPC7 [Homo sapiens]	1865	1847/1866 (98%)	1853/1866 (98%)	0.0
gi 13562153 gb AAK19738.2 AF346629.1 (AF346629)	channel-kinase 1 [Homo sapiens]	1864	1845/1866 (98%)	1851/1866 (98%)	0.0
gi 14009344 gb AAK50377.1 (AY032951)	LTRPC7 [Mus musculus]	1863	1748/1866 (93%)	1803/1866 (95%)	0.0

<u>gi 10946830 ref NP_067425.1 </u> (NM_021450)	transient receptor potential M7; transient receptor potential-related protein, ChaK [Mus musculus]	1863	1747/1866 (93%)	1803/1866 (96%)	0.0
<u>gi 14211383 gb AAK57433.1 AF376052.1</u> (AF376052)	transient receptor potential phospholipase C interacting kinase [Mus musculus]	1862	1747/1866 (93%)	1802/1866 (95%)	0.0

Table 4F-G lists the domain descriptions from DOMAIN analysis results against NOV4. This indicates that the NOV4 sequence has properties similar to those of other proteins known to contain this domain.

Table 4F. Domain Analysis of NOV4

gnl|Pfam|pfam02816, MHCK_EF2_kinase, MHCK/EF2 kinase domain family. This family is a novel family of eukaryotic protein kinase catalytic domains, which have no detectable similarity to conventional kinases. The family contains myosin heavy chain kinases and Elongation Factor-2 kinase and a bifunctional ion channel.

CD-Length = 206 residues, 94.7% aligned

Score = 79.7 bits (195), Expect = 1e-15

Table 4G. Domain Analysis of NOV4

gnl|Pfam|pfam00520, ion_trans, Ion transport protein. This family contains Sodium, Potassium, Calcium ion channels. This family is 6 transmembrane helices in which the last two helices flank a loop which determines ion selectivity. In some sub-families (e.g. Na channels) the domain is repeated four times, whereas in others (e.g. K channels) the protein forms as a tetramer in the membrane.

CD-Length = 191 residues, 99.0% aligned

Score = 62.8 bits (151), Expect = 2e-10

Capacitative calcium entry (CCE) describes Ca^{2+} influx into cells that replenishes Ca^{2+} stores emptied through the action of IP₃ and other agents. It is an essential component of cellular responses to many hormones and growth factors. The molecular basis of this form of Ca^{2+} entry is complex and may involve more than one type of channel. Studies on visual signal transduction in *Drosophila* led to the hypothesis that a protein encoded in transient receptor potential (Trp) and related proteins may be a component of CCE channels. Zhu et al.)

reported the existence of six trp-related genes in the mouse genome. Expression in L cells of small portions of these genes in antisense orientation suppressed CCE. Expression in COS cells of two full-length cDNAs encoding human trp homologs, Htrp1 and Htrp3, increased CCE. This identifies mammalian gene products that participate in CCE.

5 Human TRPC genes encode proteins with sequence similarity to the *Drosophila* 'transient receptor potential' (trp) gene product. TRPC proteins are thought to be subunits of capacitative calcium entry (CCE) channels, which mediate calcium influx into cells to replenish internal stores of calcium. Using exon trapping on a contig from 21q22.3, Kudoh et al. (1997) isolated an exon whose deduced amino acid sequence shows similarity to the
10 sequences of human TRPC and *Drosophila* trp proteins. Nagamine et al. (1998) isolated human fetal brain and caudate nucleus cDNAs corresponding to the exon and its parent gene. The deduced 1,503-amino acid protein, which is named TRPC7, is 22.9% identical to human TRPC1 (602343), 21.2% identical to human TRPC3 (602345), and 22.6% identical to *Drosophila* trp. TRPC7 contains 7 predicted membrane-spanning domains. The TRPC7 gene
15 has 32 exons spanning approximately 90 kb. Northern blot analysis of human tissues detected a 6.5-kb TRPC7 transcript predominantly in fetal and adult brains, where it was expressed in several regions. In caudate nucleus and putamen, a putative 5.5-kb alternatively spliced TRPC7 product also was detected.

The disclosed NOV4 nucleic acid of the invention encoding a TRANSIENT
20 RECEPTOR POTENTIAL-RELATED PROTEIN-like protein includes the nucleic acid whose sequence is provided in Table 4A or 4C or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 4A or 4C while still encoding a protein that maintains its TRANSIENT RECEPTOR POTENTIAL-RELATED PROTEIN-like activities and
25 physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way
30 of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or

neurodegeneration, arthritis, tendonitis, diabetes, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, cirrhosis, lymphedema, ulcers, tonsillitis, or other pathologies or conditions. The NOV4 nucleic acid encoding the TRANSIENT RECEPTOR POTENTIAL-RELATED PROTEIN-like protein of the invention, or fragments thereof, may further be
5 useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV4 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV4 substances for use in therapeutic or
10 diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV4 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology
15 of the disease and development of new drug targets for various disorders.

NOV5

A disclosed NOV5 nucleic acid of 1869 nucleotides (also referred to as CG57609-01) encoding a Epsin-3-like protein is shown in Table 5A. Putative untranslated regions upstream
20 and/or downstream from the coding region, if any, are underlined, and the start and stop codons are in bold letters.

Table 5A. NOV5 nucleotide sequence (SEQ ID NO:11).

<p> <u>GCGGGGGCGAGGGCCACCCACCTCCAAGTCTCCAGCCATGACGACCTCCGCACTCCGGCG</u> <u>CCAGGTGAAGAACATCGTGCACTACTCCGAGGCAGAAATCAAGGTGCGCGAGGCCAC</u> <u>CAGCAATGACCCCTGGGGCCCCCTAGTTCGCTCATGTCCGAGATCGCTGACCTGACCTT</u> <u>CAACACAGTGGCCTTCACCGAAGTCATGGGCATGCTGTGGCGGCGGCTCAATGACAGCGG</u> <u>CAAGAACTGGCGGCACGTGTACAAGGCTCTAACATTGCTGGACTACCTGCTCAAGACGGG</u> <u>CTCCGAGCGGGTGGCCACCAAGTGCCTGCGAGAACCTCTACACCATCCAGACACTCAAGGA</u> <u>CTTCCAGTACATCGACCGCGACGGCAAGGACCAGGGCGTCAACGTGCGCGAGAAGGTCAA</u> <u>GCAGGTGATGGCCCTGCTCAAGGATGAGGAGCGGCTGCGGCAGGAGCGAACCCACGCCCT</u> <u>CAAGACCAAGGAGCGCATGGCACTGGAGGGCATCGGCCCGCTGGTGCTGGGCTTCAGCCG</u> <u>CCGCTACGGCGAGGACTACAGCCGCTCCCGGGGCTCCCGCTCCTCCTACAACCTCCTCCTC</u> <u>TTGCTACCCCGCTATACCTCCGACCTGGAGCAGGCCCGGCTCAGACGTCAGGGGAAGA</u> <u>GGAAGTGCAGCTGCAGCTGGCCCTCGCCATGAGCCGTGAGGAGGCAGAGAAGGAGGTGAG</u> <u>GTCTTGGCAGGGTGTGGCTCCCCCATGGCCAATGGTGACAGGGGCGGTGGTCCACATCA</u> <u>GCGGGACAGAGAGCCTGAGAGAGAAGAGAGAAAGGAGGAGGAGAAGCTAAAAACCAAGCA</u> <u>GTCTCCATCCTGGACTTGGCTGACATCTTCGTACCTGCCCTGGCCCCGCCCTCCACACA</u> <u>CTGCTCTGCTGACCATGGGACATCCAGGTTTTAGGCCGAACACAGAGGCCAGTGGATC</u> <u>CTCCTGGGGGCTTCTGCAGACCCCTGGTCTCCGATCCCCCTCAGGAACCGTCTGTCCCG</u> <u>AAGCCAGCCCTGGGATCTGACTCCCATGCTCTCTCTCTGAGCCCTGGGGCAGGACCCC</u> </p>
--

AGTGCTGCCTGCTGGGCCCCCACCACAGACCCCTGGGCCCTGAACTCTCCCCACCACAA
 ACTCCCCAGCACTGGGGCTGACCCTTGGGGAGCCTCCCTGGAGACCTCCGACACACCTGG
 TGGTGCCTCGACCTTTGACCCATTTGCCAAACCTCCAGAATCCACAGAGACCAAGGAGG
 GCTGGAGCAGGCCCTGCCCTCTGGGAAGCCCAGCAGCAGCGGGGAGCTGGACCTGTTTGG
 AGACCCCAGCCCCAGTTCCAAGCAAAATGGCACGAAGGAGCCAGATGCCCTGGACCTGGG
 CATACTAGGGGAAGCACTAACCAGCCAAGCAAAGAGGCCCGAGCTTGCCGGACTCCCGA
 GTCCTTCCTGGGTCCCTCAGCTTCCTCCTTGGTCAACCTTGACTCGTTGGTCAAGGCACC
 CCAGGTTGCAAAGACCCGGAACCCCTTCCTGACAGGTGGTCTCAGCGCTCCGTCCCCCAC
 CAACCCGTTTCGGCGCGGGCGAGCCGGGCAGGCCGACGCTAAACCAGATGCGCACCGGCTC
 GCCGGCGCTGGGCCCTGGCAGGCGGGCCTGTGGGGGCGCCCTGGGCTCCATGACCTACAG
 CGCCTCTCTGCCCCCTCCGCTCAGCAGCGTGCCAGCTGGCTTGACCCCTCCCGCCCTCGGT
 TAGCGTCTTCCCGCAGGCCGAGCCTTCGCACCGCAGCCGCTGCTGCCCCACGCCGAGCTC
 AGCCGGGCGCGGCCCGCCCCCGCAGACCGGCACCAACCCCTTCCTCTGAGCCCCGCC
 CCGTCCCAT

In a search of public sequence databases, the NOV5 nucleic acid sequence, located on chromosome 17 has 1210 of 1234 bases (98%) identical to a gb:GENBANK-

ID:AK000785|acc:AK000785.1 mRNA from Homo sapiens (Homo sapiens cDNA FLJ20778

5 fis, clone COL05704). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV5 polypeptide (SEQ ID NO:12) encoded by SEQ ID NO:11 has 604 amino acid residues and is presented in Table 5B using the one-letter amino acid code.

Signal P, Psort and/or Hydropathy results predict that NOV5 has no signal peptide and is

10 likely to be localized in the cytoplasm with a certainty of 0.4500.

Table 5B. Encoded NOV5 protein sequence (SEQ ID NO:12).

MTTSALRRQVKNI VHNYSEAEIKVREATSNDPWGPPSSLMSEIADLTFNTVAFTEVMGML
 WRR LND SGKNWRHVYKALTLLDYLLKTGSEVAHQCRENLYTIQTLKDFQYIDRDGKDQG
 VNVREKVKQVMALLKDEERLRQERTHALKTKERMALEGIGPLVLGFSRRYGEDYSRSGS
 PSSYNSSSSSPRYTSDLEQARPQTSGEELQLQLALAMSREEAEKEVRSWQGDGSPMANG
 AGAVVHHQRDREPEREERKEEEKLKTSQSSILDLADIFVPALAPPSTHCSADPWDIPGFR
 PNTEASGSSWGSPADPWSPIPSGTVLSRSQPWDLTPMLSSSEPWGRTPVLPAGPPTTDPW
 ALNSPHHKLPSTGADPWGASLETSDTPGGASTFDPFAKPPESTETKEGLEQALPSGKPSS
 SGELDLFGDPSPSSKQNGTKEPDALDLGILGEALTQPSKEARACRTPESFLGPSASSLVN
 LDSLVKAPQVAKTRNPFLTGGLSAPSPTNPFAGAGEPGRPTLNQMRTGSPALGLAGGPVGA
 PLGSMTYSASLPLPLSSVPAGLTLPASVSVFPQAGAFAPQPLLPTSSAGPRPPPPQTGT
 NPFL

A search of sequence databases reveals that the NOV5 amino acid sequence has 398 of 441 amino acid residues (90%) identical to, and 406 of 441 amino acid residues (92%) similar

15 to, the 632 amino acid residue ptrn:TREMBLNEW-ACC:AAG45223 protein from Homo sapiens (Human) (EPSIN 3). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

5

10

20

30

of whose residues may be changed from the corresponding residue shown in Table 5B while still encoding a protein that maintains its Epsin-3-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 10 percent of the residues may be so changed.

5 The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

 The above defined information for this invention suggests that this Epsin-3-like protein (NOV5) may function as a member of a “Epsin-3 family”. Therefore, the NOV5 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in
10 (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing
15 (but not limited to) those defined here.

 The NOV5 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the Epsin-3-like protein (NOV5) may be useful in gene therapy, and the Epsin-3-like protein (NOV5) may be useful
20 when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from psoriasis, actinic keratosis, tuberous sclerosis, acne, hair growth/loss, alopecia, pigmentation disorders, endocrine disorders, or other pathologies or conditions. The NOV5 nucleic acid encoding the Epsin-3-like protein of the invention, or fragments thereof, may
25 further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

 NOV5 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV5 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the
30 art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV5 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV6

A disclosed NOV6 nucleic acid of 2646 nucleotides (also referred to as CG57611-01) encoding a CD22-like protein is shown in Table 6A. Putative untranslated regions upstream and/or downstream from the coding region, if any, are underlined, and the start and stop codons are in bold letters.

Table 6A. NOV6 nucleotide sequence (SEQ ID NO:13).

ATGGACAACCCACAGGCTCTGCCACTCTTCTACTCCTGGCCTCCTTGGTAGGGATCCTC
 ACCCTCAGAGCCCTCTTCTGGACTTCAGCAAACCAACTTCTCCTCTGCCTTCTCTTCAGAC
 TCAAAGAGCTCTTCCCAGGGGCTGGGTGTGGAAGTTCCCTCCATCAAACCTCCCAGCTGG
 AAAGTTCCAGATCAGTTCTGGATTCAAAGCCTCTGCTGGAATCTCTGATTCCAGCTGG
 TTTCTGAGGCCCTGAGTTCCAACATGTCTGGGTCTTCTGGTCAAATGTTTCTGCTGAG
 GGCCAAGATTGAGCCCGGTTTCCCCCTTCTCTGAAACCCCTGGTTCTGAAGTATTTCCT
 GATATTTTCGGATCCTCAAGTTCTGCCAAAGACCCCAAGCCTTCTTCTACTGTTAAGACC
 CCAGCTTCAAACATTTCTACTCAAGTCTCCCATACCAAAGTCTGTTGAGGCCCCAGAT
 TCAAATTTCTCCCCGATGATATGGATCTTAACTCTCTGCCAGAGCCCTGAATCCAAA
 TTTTCTGCAGAGACCCACTCAGCTGCAAGCTTCCCCAGCAGGTGGGGGGCCCACTCGCT
 GTGCTGGTGGGGACCACCATCCGGCTCCCCCTAGTCCCAATCCCCAACCCCTGGGCCCCC
 ACCTCTCTGGTGGTCTGGCGCCGGGGCTCAAAGGTGCTGGCAGCTGGGGGCCTGGGGCCA
 GGGGCACCTCTGATCAGCCTGGACCCTGCTCACCGAGACCACCTGCGATTGACCAGGCC
 CGGGGGGTTCTGGAGCTCGCCTCTGCCAGCTGGACGATGCAGGGGTCTACACGGCTGAG
 GTCATCCGGGACAGGGTCTCCAGCAGACTCACGAGTTCACGGTGGGTGTGTATGAGCCC
 CTACCCAGCTGTCTGGTTTCAGCCCAAGGCTCCAGAGACAGAGGAGGGGGCGGCCGAGCTC
 CGGCTGCGCTGCTGGGGTGGGGGCCAGGTGCGGGGAGCTGAGCTGGAGCCGGGACGGA
 CGCGCCCTGGAGGCGGCGGAATCGGAGGGAGCCGAGACGCCCCGGATGCGCTCAGAGGGC
 GACCAGCTGCTCATCTGCGCCCTGTGCGCAGCGACCACGCCCCGTACACTTGCCGCGTC
 CGCAGCCCCCTTCGGCCACAGGGAGGCTGCCGCCGACGTCAGCGTCTTCTACGGCCCGGAC
 CCGCCGACCATCACGGTCTCCTCGGACCGCGACGCGCGCCTGCCCGCTTTGTACCCGCG
 GGCAGTAACGTGACCTTGCCTGCGCTGCGCCGCGCCTCGCGGCCGCCCGGACATCACGTGG
 AGCCTGGCGGACCCGCGCCGAGCCCGCGGTGCGCGCGGGGTGCGCCTCTGTCGCCGCG
 GTCGGACCGGGCCACGCAGGCACCTACGCCCTGCTGGCGGCGAACC CGGTACCGGCCGC
 CGCCGCCGCTCGCTGCTCAACCTTACAGTGGCGGACCTGCCCCCGGGGGCCCCACAGTGC
 TCAGTTGAAGGGGTCCCGGGGACCGCAGCCTCCGCTTCCGCTGCTCGTGGCCCGCGGG
 GCCCCGTGCTGCCCTCCCTGCAGTTCCAGGGTCTCCCGAAGGCATCCGCGCCGGGCCAGTG
 TCCTCTGTGCTGCTGGCGGCCGTCCCCGCCACCCCCGGCTCAGCGGCGTCCCCATCACC
 TGCCTTGCTCGCCACCTGGTGGCCACGCGTACCTGCACAGTCACGCCGAGGCCCCCGGA
 GAGGTGCTGCTGCATCCGCTGGTGGCAGAGACAGGTTGGGGGAGGCAGAGGTGGCACTG
 GAGGCCTCTGGTTGTCCCCCACCTCACGGGCATCCTGGGCCCGGGAAGGGAGGCCCTG
 GCTCCAGGAGGCGGGAGTCGCCTGCGGCTCAGTCAAGATGGGCGGAAACTCCACATCGGC
 AACTTCAGCCTGGATTGGGACCTGGGAAATTACTCCGTGCTGTGAGTGGGGCGCTGGGT
 GCTGGCGGTGACCAGATCACCTCATTTGGACCTCCATATCCTCGTGGAGGCTTCAGAGA
 GCCAGAGATGCACCGTGCTGACTTGGGATGTGGAGCGCGGGGGCCCTGATCAGCAGTTTT
 GAGATCCAGGCATGGCCAGATGGGCCTGCTCTGGGCAGGACTTCCACCTACAGGGACTGG
 GTCTCCCTGCTCATCTGGGCCCTCAGGAGCGGTGAGCCGTGGTGGCCCTTCCACCTCGG
 AACCAGGGACCTGGACCTTTTCGATCCTGCCATCCTGGGGGGCCAGCCAGGGACTCCA
 TCACAAAGCCGGTCTACCGGGCGGCCCCACGTTGAGCCATGGGGCCATCGCTGGCATC
 GTCCTGGGCTCCCTGCTGGGCCTGGCGCTGCTAGCCGTAATTCTCCTCCTTTGCATCTGC
 TGCCTGTGCCGCTTTCGTGGAAGACTCCTGAGAAAAAGAAGCATCCTTCTACCTTGGTC
 CCCGTGGTCACCCCCTCAGAAAAAGAAGATGCATAGTGTGACCCAGTGGAGATTTTCATGG
 CCTCTGGACCTCAAAGTCCCTCTGGAGGACCACAGCTCAACTAGGGCCTACCAAAAAGAAG
 AGTCTTCCCTGTATTTGTGCAAGTAAGGAGATGTGACTTCTTGGCTGGGAAACTCTTGCTG
 ATCTAA

5

10

Table 6B. Encoded NOV6 protein sequence (SEQ ID NO:14).
MDNPQALPLFLLLASLVGILTLRASSGLQQTNFSSAFSSDSKSSSQGLGVEVPSIKPPSW KVPDQFLDSKASAGISDSSWFPEALSSNMSGSFWSNVSAEQDLSPVSPFSETPGSEVFP DISDPQVPAPKDPKPSFTVKTPASNISTQVSHTKLSVEAPDSKFSRDDMDLKLKLSAQSPESK FSAETHSAASFPQQVGGLPAVLVGTITIRLPLVPIPNPGPPTSLVVWRRGSKVLAAGGLGP GAPLISLDPAHRDHLRFDAQRGVLELASAQLDDAGVYTAEVIRAGVSQQTHEFTVGVEYP LPQLSVQPKAPETEEGAELRLCLGWGPGRGELSSVTRDGRALEAAESEGAETPRMRSEG DQLLIVRPVPSRDHARYTCRVRSFPGHREAAADVSVFYGDPPTITVSSDRDAAPARFVTA GSNVTLRCAAASRPADITWSLADPAEAAVPAGSRLLLPVAVGPGHAGTYACLAANPRTGR RRRSLNLTVADLPFGAPQCSVEGGPGDRSLRFRCSWPGGAPAAASLQFQGLPEGIRAGPV SSVLLAAVPAHPRLSGVPIITCLARHLVATRTCTVTPEAPREVLLHPLVAETRLGEAEVAL EASGCPPPSRASWAREGRPLAPGGGSRLRLSQDGRKLHIGNFSLDWDLDGNYSVLCSGALG AGGDQITLIGPSISSWRLQRADAAVLTWDVERGALISSFEIQAWPDGPALGRSTSTYRDW VSLILIGPQERSAVVPLPPRNPGTWTFRILPILGGPQPTPSQSRVYRAGPTLSHGAIAGI VLGSLGLALLAVLLLLCICCLCRFRGKTPEKKKHPSTLVPVVTPEKKMHSVTPVEISW PLDLKVPLEDHSSTRAYQKKSPLPVFVQVRRCDFLAGKLLLI

15

20

Table 6C. BLAST results for NOV6

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 312584 emb CAA47697.1 (X67280)	biliary glycoprotein [Mus musculus]	458	68/237 (28%)	109/237 (45%)	1e-16
gi 14029256 gb AAK52602.1 (AF287912)	CEA-related cell adhesion molecule 2 [Mus musculus]	520	68/237 (28%)	109/237 (45%)	3e-16
gi 423398 pir S34338	biliary glycoprotein F - mouse	521	68/237 (28%)	109/237 (45%)	4e-16
gi 483309 pir JC1509	biliary glycoprotein E - mouse	458	68/237 (28%)	108/237 (44%)	5e-16
gi 109630 pir S11626	carcinoembryonic antigen - mouse (fragment)	379	66/237 (27%)	106/237 (43%)	1e-15

Table 6D lists the domain descriptions from DOMAIN analysis results against NOV6. This indicates that the NOV6 sequence has properties similar to those of other proteins known to contain this domain.

Table 6D. Domain Analysis of NOV6

[gnl|Smart|smart00408](#), IGc2, Immunoglobulin C-2 Type
CD-Length = 63 residues, 93.7% aligned
Score = 55.5 bits (132), Expect = 1e-08

5

The disclosed NOV6 novel gene described here contains three immunoglobulin domains and has homology to mouse CD22, a B lymphocyte-restricted adhesion molecule, mouse colon biliary glycoprotein, and carcinoembryonic antigen. The immunoglobulin domain is found as a tandem repeat in Streptococcal cell surface proteins, such as the IgG binding proteins G and MIG. These proteins are type I membrane proteins that bind to the constant Fc region of IgG with high affinity. The N-terminus of MIG mediates binding to plasma proteinase inhibitor alpha 2-macroglobulin after complex formation with proteases.

The human B lymphocyte-specific Ag, CD22, is a cell adhesion molecule expressed on the surface during a narrow window of B cell development, coincident with surface IgD. A ligand for CD22 has recently been identified on human T cells as the low molecular mass isoform of the leukocyte common Ag, CD45RO. CD22 has been reported to function in the regulation of both T and B cell activation in vitro.

Carcinoembryonic antigen (CEA) is a widely used tumor marker, especially in the surveillance of colonic cancer patients. Although CEA is also present in some normal tissues,

it is apparently expressed at higher levels in tumorous tissues than in corresponding normal tissues. Carcinoembryonic antigen (CEA) expression is perhaps the most prevalent of phenotypic changes observed in human cancer cells. Twenty-seven CEA cDNA clones were isolated from a human colon adenocarcinoma cell line. Most of these clones are full length and
 5 consist of a number (usually three) of surprisingly similar long (534 base pairs) repeats between a 5' end of 520 base pairs and a 3' end with three different termination points. The predicted translation product of these clones consists of a processed signal sequence of 34 amino acids, an amino-terminal sequence of 107 amino acids, which includes the known terminal amino acid sequence of CEA, three repeated domains of 178 amino acids each, and a
 10 membrane-anchoring domain of 27 amino acids, giving a total of 702 amino acids and a molecular weight of 72,813 for the mature protein. The repeated domains have conserved features, including the first 67 amino acids at their N termini and the presence of four cysteine residues. Comparisons with the amino acid sequences of other proteins reveals homology of the repeats with various members of the immunoglobulin supergene family, particularly the
 15 human T-cell receptor gamma chain. CEA cDNA clones in the SP-65 vector were shown to produce transcripts in vitro which could be translated in vitro to yield a protein of molecular weight 73,000 which in turn could be precipitated with CEA-specific antibodies (See Schrewe H et al., Mol Cell Biol 1990 Jun;10(6):2738-48.).

The biliary glycoprotein (BGP)-encoding gene is a member of the human
 20 carcinoembryonic antigen (CEA) gene family. McCuaig et al. cloned several mouse Bgp cDNAs from an outbred CDR-1 mouse colon cDNA library, as well as by reverse transcription-PCR amplification of colon RNA. The distinguishing features of the deduced Bgp protein isoforms are found in the two divergent N-terminal domains, the highly conserved internal C2-set immunoglobulin domains, and an intracytoplasmic domain of either 10 or 73
 25 amino acids (aa). The cDNA structures suggest that these mRNAs are produced through alternative splicing of a Bgp gene and the usage of multiple transcriptional terminators. The Bgp deduced aa sequences are highly homologous to several well characterized rat hepatocyte proteins such as the cell CAM105/ecto-ATPase/pp120/HA4 proteins. Oligodeoxyribonucleotide probes representing the various cDNA isoform domains revealed
 30 predominant transcripts of 1.8, 3.1 and 4.0 kb on Northern analyses of mouse colon RNA; some of these bands are actually composed of several co-migrating transcripts. The transcripts encoding the long intracytoplasmic-tailed Bgp proteins are expressed at one-tenth the relative abundance of the shorter-tailed species. The expression of the many Bgp isoforms at the surface of epithelial cells, such as colon, suggests that these proteins play a determinant role,

through self- or heterologous contact, in renewal and/or differentiation of their epithelia (See McCuaig et al., Gene 1993 May 30;127(2):173-83).

The disclosed NOV6 nucleic acid of the invention encoding a CD22-like protein includes the nucleic acid whose sequence is provided in Table 6A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 6A while still encoding a protein that maintains its CD22-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 41 percent of the bases may be so changed.

The disclosed NOV6 protein of the invention includes the CD22-like protein whose sequence is provided in Table 6B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 6B while still encoding a protein that maintains its CD22-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 71 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this CD22-like protein (NOV6) may function as a member of a "CD22 family". Therefore, the NOV6 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV6 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the CD22-like protein (NOV6) may be useful in gene therapy, and the CD22-like protein (NOV6) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from endometriosis, fertility, systemic lupus erythematosus, autoimmune disease, asthma, emphysema, scleroderma, allergy, ARDS, or other pathologies or conditions. The NOV6 nucleic acid encoding the CD22-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV6 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV6 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV6 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV7

A disclosed NOV7 nucleic acid of 8589 nucleotides (also referred to as CG57595-01) encoding a MEGF8-like protein is shown in Table 7A. Putative untranslated regions upstream and/or downstream from the coding region, if any, are underlined, and the start and stop codons are in bold letters.

Table 7A. NOV7 nucleotide sequence (SEQ ID NO:15).

TGCAGGAGGCGGCGATGGCCCTGGGCAAGGTTCTGGCCATGGCACTGGTTTTGGCCCTGGCCCGTGCTGGG
 GTCGCTGTCCCCTGGGGCCCGGGCGGGGACTGCAAGGGGCAGCGGCAGGTGCTGCGGGAGGCGCCAGGC
 TTCGTGACGGATGGTGGGGCAACTACAGCGTCAATGGCAACTGCGAGTGGCTCATCGAGGCCCAAGCC
 CCCAGCACCGGATCCTGCTGGACTTCCTTTTCTGGACACAGAGTGACGTATGACTACCTGTTTCGTGTA
 TGACGGTGACTCCCGCGAGGGCCGCTGCTTGCCAGTCTAAGTGGGAGCACCCGACCTCCGCCCATCGAA
 GCTTCCTCAGGCAAGATGCTGCTGCACCTCTTCAGTGATGCCAACTACAACCTGCTGGGCTTTAACGCCT
 CATTCGCTTCTCCCTGTGCCCGGGTGGCTGCCAGAGCCACGGGCAGTGCCAGCCACCGGGTGTGTGTC
 CTGCGAGCCGGGCTGGGGGGTCTGACTGTGGCCTGCAGGAGTGCTCAGCCTACTGTGGCAGCCACGGC
 ACCTGGCCTCGCCCCTGGGACCATGCCGCTGTGAGCCTGGCTTCTTGGGACGTGCCTGTGACCTGCACC
 TGTGGGAGAACCAGGGGGCTGGGTGGTGGCACAACGTGAGTGCCAGGGACCTGCCTTCTCTGCCGTAT
 TGGGGCAGCTGGCGCCTTCTGTCCCCACAGGGCTGCTGGCAGTTTTCGGAGGCCAGGACCTCAACAAT
 GCCCTGGGTGACCTCGTCTATACAACCTTCCGCCAACACCTGGGAGTCTTGGGACCTGAGTCTCTGCC
 CGGCTGCCCGTCACTCCCATGTGGCCGTGGCCTGGGCGGCTCCCTGGTACTGATGGGTGGTGGTGGC
 TGACGGCTCGCTACCAACGACGTGTGGGCCTTCAGTCCACTGGGCAGGGGCCACTGGGAGCTCCTGGCA

CCACCTGCCTCCAGCTCCTCGGGGCCCCAGGCTGGCAGGTACGCGGCTGCCCTGGTGGATGATGTCT
GGCTATATGTGTCTGGAGGCCGACCCCGCAGACCTCTTCTCCTCTGGCCTCTTCCGTTTCCGCCTTGA
CAGCACAGCGGGGGCTATTGGGAGCAGGTGATTCGGCAGGCGGACGGCCCCCTGTGCCACTGGCCAC
TCCATGGTGTTCATGCCCCCTCCCGTGCCTGTGGTCCATGGTGGACACCGGCCCTCCACTGCCCGGT
TCTCTGTGCGAGTGAACCTCACTGAGCTTTTCCACGTGGATCGGCATGTGTGGACGACGCTGAAGGGGCG
GGATGGGCTTCAGGGCCCCAAGGAGCGAGCCTTCCACACAGCCAGTGTTCGGGCAATTACATGGTGGTC
TATGGGGGCAATGTGCACACCCATTACCAGGAGGAAAAGTGCTACGAAGATGGCATCTTCTTACCACC
TTGGCTGCCATCAATGGGTGTGAGGCTGAGCTTGCCCGCCAGGAACCCCTGAGGGCCGAGCAGCGCC
TCCAGTGGTGGTACTCACATGTAGCTGCGGTGCTTGGTGGCAGCGTCTGTGGTGGCTGGGGGTAC
AGCGGCCGCCCCGTGGGGACTTGATGGCGTACAAGGTGCCCCCTTTGTGTTCCAGGCACCTGCCCTG
ACTACCACTTGGACTACTGTCCATGTACACAGACCACAGCGTCTGCTCCCGGACCCGGAATGCAGTTG
GTGCCAAGGAGCCTGCCAAGTGCACCCCTCTGGGACCCCTTGGGGGCTTGTCCAGCCGCCAGCTGC
CTGGGCTTGGGCCGCTCTGGGTGACTGCCAGGCTGCTGGCCTTCCAGCAGCCCCACAGCCCTCCAC
GGGACCTTGGCACCTGGGCTGGTGGCTGCACAATGAGAGCTGCCTCCCTAGGCCTGAGCAGGCCGCTG
CCGAGGGGAGCAGATCTCAGGCACCTGTGGGCTGGTGGGGGCTGCGCCTGTCTTCGTACAGTCCCTGGAG
GCCTGCGTCACCCAGAGCTTCTGCTTGGCCTGCACCTTGCTCACCTTCCAGCAGCCGCCAATACCTCCC
AGCTTGACAAGGTCTCAATTGTCCGAGCAGACCACTACCCCTAACACCCAGCGCAGAGACAGATGTGT
CCTGGTCTACCGTGGCTTCTATCTACCAATGTGCTGCTGGAGGGCCAGGTGGACCAGGGGCTGAGGACGTG
GCCGTGTGGACGCGGGCCAGCGCCTACACGTCTGCGCCGGATGGCCCGTGGCCCTGACACGGAGAACA
TGGAGGAGGTGGGGCGCTGGGTGGCTCATCAGGAGAAGGAGACGCGGGGCTGACGCGCCCTGGGTCTGC
TCGCTGTTCCTCTGCTTGGCGGGACCAAGTATGCAGTAGAGATCCAGGGCCAGCTCAATGGCTCG
GCAGGCCCTGGGCACAGCGAGCTAACTCTGCTGTGGGATCGGACTGGTGTGCCAGGAGGCAGCGAGATCT
CCTTCTTCTCTCTGGAGCCCTACCGCTCGTCTGCTGACCTCTTATTCTTCTGCTGGGCTGCTTGGC
AGACCAGGGCTGTGGCTGGTGCCTGACAGTGCCACCTGCCACCTGCGCCAGGGGCGAGCCATTGCGGG
GATGACGGGCTGGTGGGTCCCTGTGGTGTGGTGCCTTACCCTTGGCCACTCTGCGAGGACATCGGCTCGG
ACTGCCACGCTGCACCCAGGACCCCTTCTGTGAGTGGCATCAGAGCACCAGCCGAAAGGGGACGCGGC
ATGCAGCCGCGGGGCGGGGTGGGGTGCCTGAAAGATCCAGAGGAGTGTCCCGCTCTGCGAGCCAG
CGACTGACCTGTGAGGACTGCTTGGCCAACTTAGCCAGTGCGCTGGTGGCAGTCCACCCACACTGTCT
TCCTGTTTGTGCTTACTTGGCCCGGTACCCACAGGGGCTGTGAGGCTGGGACGACAGTGTACACTC
GGAGCCACGGTGCCGAGCTGCGATGGCTTCTGACCTGCCATGAGTGTCTGACAGCCACGAGTGTGGC
TGGTGTGGCAATGAGGACAACCCCACTGGGACGGTGCCTACAGGGGACTTCTCAGGGCCCTCGGTG
GGGGTAATGCTCCTGTGGGTGGGGGAGGGCTGGGGCTTCCCGTGGCCCTCCCTGCCCGCTGGGCATA
GCCCGCTGTCTGACGTGGATGAGTGTGCTGGGCTGGCCCGTGGCCCGGCGGACCTGACCTGCAACC
AACCGTCTCAGTACGAGTGTCACTGCGAGCGGGCTACAGGGTGTGGCATCTCACACTGTCAACC
GCACGGATGGCATCTCACACTGCAACCGCACGTGCTTGGAGGACTGTGGCCATGGTGTGTGACGTGGCCC
CCCGACTTTACCTGCGTGTGTGACCTAGGCTGGACATCAGACCTGCCCCCTCCACACCCGCCCCGGGT
CCGCCAGCCCCCGCTGTCTCCGGGACTGTGGCTGCAGCTTCCACAGCCACTGCCGCAAGCGGGGCTGTG
GCTTCTGCGACAGTGCAGGACTGGACATGGGGGAGCACTGCGAACGATGCGGGGCGGAGCTTGG
CAACGCCACAGGCTCTAGGGGCTGCCGCCCTGCCAGTGAACGGGACGCGGACCCAGCCGTGGCCAC
TGGGACAACCTCAGTGGGCTGTGCTTCTGCCAGGACCAACCGAGGGTGCCCACTGCCAGCTGTGCTCCC
CAGGCTATTATGGGGATCCCCGGGCGGGTGGTTCCTGCTTTCGGGAGTGTGGGGGTGCGGCCCTCCTCAC
CAACGTGCTCAGTGGCACTGGGCTCACGCCGGTGGGGGCTGCTGCTTCCAGGTGCGGGGCTGCA
AGAGCCGGGCTGGCTGTCTACTGTGTGGGTGTCTCGGCCACTGAGGAGCTACAGCCCTGTGCTC
CCGGGACCTCTGTCCCCCACTCACCTCACCTTCTCCCCGACAGCAGCAGCCCTGACAGCTGAGCTA
CGTCTGGCGTTTATGATGATTTCCACGCTTCTTGGACACTGGTGTGTGTCAGTGGACCGCAGCCTCATA
GCTGCTTCTGCGGCCAGCGAGGACAGGCCCTCACTGTTCAGGCCCTGTCTGGGCTGTCTGCTGCTG
ACTGGGAGGCCAATGGCTCCTCATCTGGGGCTTCAATGCTTGGTGGGCTCTGCCCGCTGTGGGTGAGG
GGGCCCCGGGAGCTGTCCCGTCCCCAGGAATGCGTGCCCGAGGACGGTGTGACAGTGGCGGGCTCTGC
CGATGTCTCAGGGCTGGGCTGGGCCACACTGCCGATGGCTCTGTGCTCTGAGAACTGCAATGCCACACA
CTGGGGCAGGAATTGTAAACAGAGCCTGGGTGTGTGCATCTGTGCCGAGGCTTCCGGGGCCCCGACTG
CGCCACCAAGCTGGATGGCGGGCAGCTGGTGTGGGAGACCTCATGGACAGCCGCTCTCAGCCGACACT
GCCAGCCGCTTCTGCACCCGCTGGGCCACACCATGGTGGATGGACCCGATGCCACCTTGTGGATGTTTG
GGGGCTGGGCTGCCCCAGGGGCTGTGGGAAACCTGTACAGGTAAGTGTGAGTGGCGGGTGGAC
ACAGATGTGGCGGGAGCCGAGGACGGGGGCCAGGCCCATCGCCCCGCTCCTTCCATGACGCGCATAT
GTGCCCGCTGGCCGTGGTGGCATGTATCTGCTGGGGGACTTACCGCTGGAGGCGTCAACCGTGATTCT
GGGTCTCAACCTCACACCTGCAATGGCGGAGGAGAAGGCCCCCCAGACCGTGGAGCTGCCAGCCGT
TGCTGGTCAACCCCTTACTGCCCGCCGAGGCTGTCTGCTCTGCTGGTGGGCGGTACTCCCCGAAAAT
GGCTTCAACAGCAGCTGCTGGAGTACAGCTGGCAACCGGCACCTGGGTGTGAGGAGCCAGAGTGGGA
CAGCCCGCAGAGTCTCTATGGTCACTGTGTCTACACAGGAGCCAGGCTCCCTCTACGTGTTTGG
GGGGTTCGATTCCATGTGGAGCTGGCGGCCCATCCCCGAGCTCTACTCCCTGCACTGTCTGACCGC
ACCTGGAGTCTGTGGCCCCCTTCTCAGGGGGCAAAGCGAGATCGTATGAGGAATGTGCGTGGCTCATCTC
GGGGTCTGGGCCAAGTTCTTGGGAGCAGCCTGGGTCTAGGGGGTTCGGGAAGTCAAGGAAGAAGTGGC
TCTGTGGGCTGTCTGTGGTACAGGAGTTTCTGGAGGAATCTCACCTCACCTGAAGGAGCCCGC
CCCGGCTTTTCCAGCCTCAGCCCTGTAGGGGACACCATGGTGGTCTTGGGGGCGCTCGGACCCCTG
ACGAGTTCAGCAGCGAGCTTCTGCTCTACAGGTCAACTGCAATGCCTGGCTTCTGCCGACCTCACCCG
CTGGCCTCTGTGGGGCCCCAATGGAGGAGTGTGGGCCATGTGTGGCAGCAGTGGGAGCCGCTG
TATATCTCTGGGGTTTTCGGGGAGTGGCCCTGGGCCGCTGTGGCACTGACCTGCCCCCTGACCCCT
GCCCGCTGTCTCTCACCTGAAGCTTGAACAGTCTGGGGGCTGCACCTGGTGGCATGGGGCTGTCT
GTCCGGGATCAGGCCACAGGCTGGGCTGCGGGGCTCCCCCTGCTCCCCAATGCCTCGCTCCCCGAG

GAATGTCGACGTCTCCGGACCTGCAGTGAGTGCCTGGCCCCCATCCTCGGACCCCTGCAACCTGGAGATG
GAGAGGCGTCCACCCCGCTGTAAGTGGTGTACCAACTGCCCGAAGGTGCTTGCATTGGACGCAATGG
GTCTGCGACCTCTGAGAATGACTGTGCGATCAACCAGCGAGAGGTCTTCTGGGCAGGGAACCTGCTCCGAG
GCTGCGTGCAGGGGCTGCTGACTGCGAGCAGTGCACGCGGGAGGGCAAGTGCATGTGGACGCGGCAGTTCA
AGAGGACAGGGGAGACCCGCGCATCTCTCCGTGCAGCCACCTATGACTGGACGTGCTTCAAGCCACTC
TCTGCTGAATGTGTCCCCATGCCGGTGAATCATACCCCACTGCCCTGCCCAACCCCTTGTACCTC
CTACCCCACTGTACCTCCTGCTGGACTCTAAGGAGCAGATGGGGGCTGGCAGCACTGTGTTGGAGCA
GCAGCTGCAGCAGTGTCTGAGCCCTTCTACCTGCCCTGCGATGTATGGCCGAGGGCTGTGGGCGGCT
GCTCCGGGACCTGAGAGCTGCTCCCTGGGCTGTGCTCAGGCAACTCAGTGCCTTGTGCTGCGGCGC
CCCCATGCGGCTGGTGTGCTGGGGGGCCAGGATGGGGGTGGCCGTGCTGAGGGGTGGACTCAGCG
GCCCCGTGATGGGCTGACATGTGGGCGTCCGGGGCTCCTGGGCTTCTGTCTGCCCCCTGAGGA
CGAGTGTGCAAACGGGCACACGACTGCAACGAGACGAGAATTGCCAGACAGCCCCACGGCTATGAG
TGCAGCTGCAAGACGGCTATACCATGGACAACATGACAGGGCTGTGCCGCCCTGTGTGCGCCAGGGCT
GCGTGAACGGCTCATGTGTGGAGCCCGACCACTGCCGCTGCCACTTTGGCTTTGTGGGCGCAACTGCTC
CACGAATGCCGCTGCAACCGCCACAGTGAATGCGCTGGTGTGGGGCGCGTGACCACTGCTTGTCTCTG
CGCAACCAACCAAGGGCAGCCACTGTGAGCAGTGCCTCCGCTGTTGTGGGTTCAGCTGTGCGAGGCG
GGACCTGCCGGCCCTGCCACGCCTTTTGTGCTGGAAATAGCCACATCTGCATCTCCAGGAAGGAGTTACA
AATGTCCAAGGAGAGCCAAAGAAGTACTCACTGGACCCAGAGGAGATTGAAATGAGGTGACAGAGGGT
CCTAGTGAAGACGAGGCCGTGTGCGTGAACCTGCCAGAATAACAGCTATGGGGAGAAATGCGAGAGCTGCC
TGCAGGGCTACTTCTCCTGGACGGGAAGTGACCAAAATGCCAGTGAATGGCCACGCGGACACATGTAA
CGACAGGATGGGACGGGCTGTCCATGTGAGAATAACACAGAGACGGGCACATGCCAGGGCAGCTCCCC
AGTGACCGTGCAGACTGCTACAAGTACCAGTGCGCCAAGTGCCGGGAATCATTTACGGGAGTCCGCTGG
GCGGCCAGCAGTGTACCGCCTCATCTCGGTGGAGCAGGAGTGTGCTGCGACCCACGTCACAGACCAA
TGTCTTCCATGAGCCCAACGCGGGCGCTAGGCCCGGGCGCACTGTCTCTTTGGCGTGCAGCCAAA
TTCACCAACGTGGACATCCGCTGACGCTGGACGTGACCTTCGGGGCGGTGGACCTCTATGTCTCCACCT
CCTATGACACCTTCGTGGTCCGTGTGGCCCTGACACTGGCGTCCATACTGTACACATCCAGCCACCCCC
AGCCCCACCACCTCCACCACCCCTGAGATGTTGGGCCCCGGGGGCTGGGGATCCAGGAGGAGCAGGG
GCCAGCAGTGGCCGGGCGCCCCAGCAGAGCCACGGGTACGGGAGGTATGGCCCGGGGGCTGATTACCT
ACGTGACGGTGACGGAGCCGTGCGCAGTGTGGTGGTCCGCGGCGTGCAGGACCGGCTGGTTCATCACCTA
CCACACGAGCACCATGCCCTCAAGTCGAGCCGCTTCTACCTGCTGCTGCTGGGCGTGGGAGACCCAAGT
GGGCCCGGCGCAACGGCTCAGCCGACTCGCAGGGCCTGCTCTTCTCCGGCAGGACCAGGCCACATTG
ACCTGTTTGTCTTCTTCTCGTCTTCTTCTCCTGCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTG
GAAGGCCAAGCAGGCTCTGGACAGCGGACAGGAGCAGCGCCGGCACTTGCAGGAGATGACCAAGATGGCC
AGCGCCCCCTTCGCCAAGGTACCGTCTGCTTCCACCTGACCTACTGCCCGGCTCCGCTTGGAAAGC
CGGTGGGCTCCACCTCCGCTTCCGCGCTCTGAGCCCTTCTTGGCACCCCTGCTGTGACAGGGGC
CGGTGGGCTCCGAGCCATGGGAGGGGGCTGCTGCCACCAGCCATCCCGCCACCACTGCTGGGCTG
CGAGCTGGGCCATCACTCTCGAGCCACAGAGATGGCATGGCTGGCGTGGCCACACTGCTGTCTCCAGC
TGCTGGCGGGCCCCATGCACCAACGGCGCTGCTGGGGTCAAGCCCTGTCACACTGCGGCACAGGCT
GCACGAGTACTGTGGGGTGGTGGGGTGTGCTGGGGCAGTGGGCATGGGACTGGTGCGGGCGGAAGGGA
CTGTTGAGCCAGGACAACCTCACCAGCATGTCCCTCTGACATGCCAGG

In a search of public sequence databases, the NOV7 nucleic acid sequence, located on chromosome 19 has 5224 of 5224 bases (100%) identical to a gb:GENBANK-
ID:AB011541|acc:AB011541.1 mRNA from Homo sapiens (Homo sapiens mRNA for
MEGF8, partial cds). Public nucleotide databases include all GenBank databases and the
GeneSeq patent database.

The disclosed NOV7 polypeptide (SEQ ID NO:16) encoded by SEQ ID NO:15 has 2854 amino acid residues and is presented in Table 7B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV7 has a signal peptide and is likely
to be localized at the plasma membrane with a certainty of 0.4600. The most likely cleavage
site for a NOV7 peptide is between amino acids 27 and 28.

Table 7B. Encoded NOV7 protein sequence (SEQ ID NO:16).
MALGKVLAMALVLAVALGSLSPGARAGDCKGQRQVLREAPGFVTDGAGNYSVNGNCEWL IEAPSPQHRILLDFLFDTECTYDYLFDVYDGDSPRGPLLASLSGSTRPPPIEASSGKMLL

invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV7 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 7C.

5

Table 7C. BLAST results for NOV7					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 7513144 pir T00209	MEGF8 protein - human (fragment)	1737	1737/1737 (100%)	1737/1737 (100%)	0.0
gi 14756108 ref XP_029883.1 (XM_029883)	EGF-like-domain, multiple 4 [Homo sapiens]	1214	1171/1173 (99%)	1171/1173 (99%)	0.0
gi 6681364 dbj BAA88689.1 (AB011534)	MEGF8 [Rattus norvegicus]	874	836/874 (95%)	850/874 (96%)	0.0
gi 10728654 gb AAF52597.2 (AE003619)	CG7466 gene product [Drosophila melanogaster]	2820	786/2379 (33%)	1107/2379 (46%)	0.0
gi 17862106 gb AAL39530.1 (AY069385)	LD09511p [Drosophila melanogaster]	779	192/466 (41%)	263/466 (56%)	3e-94

Tables 7D-F list the domain descriptions from DOMAIN analysis results against NOV7. This indicates that the NOV7 sequence has properties similar to those of other proteins known to contain this domain.

10

Table 7D. Domain Analysis of NOV7
<p>gnl Smart smart00042, CUB, Domain first found in C1r, C1s, uEGF, and bone morphogenetic protein.; This domain is found mostly among developmentally-regulated proteins. Spermadhesins contain only this domain.</p> <p>CD-Length = 114 residues, 82.5% aligned</p> <p>Score = 84.0 bits (206), Expect = 1e-16</p>

Table 7E. Domain Analysis of NOV7

gnl|Pfam|pfam00053, laminin_EGF, Laminin EGF-like (Domains III and V). This family is like pfam00008 but has 8 conserved cysteines instead of 6.

CD-Length = 49 residues, 87.8% aligned

Score = 47.4 bits (111), Expect = 1e-05

Table 7F. Domain Analysis of NOV7

gnl|Smart|smart00179, EGF_CA, Calcium-binding EGF-like domain

CD-Length = 41 residues, 90.2% aligned

Score = 38.5 bits (88), Expect = 0.005

The domain that characterizes epidermal growth factor (EGF) consists of
 5 approximately 50 amino acids, and has been shown to be present in a more or less conserved
 form in a large number of other, mostly animal proteins. EGF-like domains are believed to
 play a critical role in a number of extracellular events, including cell adhesion and receptor-
 ligand interactions. Proteins with EGF-like domains often consist of more than 1,000 amino
 acids, have multiple copies of the EGF-like domain, and contain additional domains known to
 10 be involved in specific protein-protein interactions. The list of proteins currently known to
 contain one or more copies of an EGF-like pattern is large and varied. The functional
 significance of EGF domains in what appear to be unrelated proteins is not yet clear. However,
 a common feature is that these repeats are found in the extracellular domain of membrane-
 bound proteins or in proteins known to be secreted (exception: prostaglandin G/H synthase).
 15 The EGF domain includes six cysteine residues which have been shown (in EGF) to be
 involved in 3 disulfide bonds. The main structure is a two-stranded beta-sheet followed by a
 loop to a C-terminal short two-stranded sheet. Subdomains between the conserved cysteines
 vary in length.

To identify proteins containing EGF-like domains, Nakayama et al. (1998) searched a
 20 database of long cDNA sequences randomly selected from a human brain cDNA library for
 those that encode an EGF-like motif. They identified several partial cDNAs encoding novel
 proteins with multiple EGF-like domains, such as EGFL4, which they named MEGF8. The
 predicted partial EGFL4 protein has a laminin-type EGF-like domain, 5 EGF-like domains,
 and a transmembrane domain. Using a radiation hybrid mapping panel, Nakayama et al.
 25 (1998) mapped the EGFL4 gene to 19q12.

The disclosed NOV7 nucleic acid of the invention encoding a MEGF8-like protein includes the nucleic acid whose sequence is provided in Table 7A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 7A while still encoding a protein that maintains its MEGF8-like activities and physiological functions, or a fragment of such a nucleic acid.

5 The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications

10 include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 0 percent of the

15 bases may be so changed.

The disclosed NOV7 protein of the invention includes the MEGF8-like protein whose sequence is provided in Table 7B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 7B while still encoding a protein that maintains its MEGF8-like activities and physiological functions,

20 or a functional fragment thereof. In the mutant or variant protein, up to about 0 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this MEGF8-like

25 protein (NOV7) may function as a member of a “MEGF8 family”. Therefore, the NOV7 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug

30 targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV7 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies

and disorders as indicated below. For example, a cDNA encoding the MEGF8-like protein (NOV7) may be useful in gene therapy, and the MEGF8-like protein (NOV7) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, multiple sclerosis, myasthenia gravis, systemic lupus erythematosus, autoimmune disease, asthma, emphysema, scleroderma, allergy, ARDS, diabetes, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, renal tubular acidosis, IgA nephropathy, or other pathologies or conditions. The NOV7 nucleic acid encoding the MEGF8-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV7 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV7 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV7 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV8

NOV8 includes two protocadherin-like proteins disclosed below. The disclosed sequences have been named NOV8a and NOV8b.

NOV8a

A disclosed NOV8a nucleic acid of 6006 nucleotides (also referred to as CG57542-01) encoding a protocadherin-like protein is shown in Table 8A. Putative untranslated regions upstream and/or downstream from the coding region, if any, are underlined, and the start and stop codons are in bold letters.

Table 8A. NOV8a nucleotide sequence (SEQ ID NO:17).

CAAATGTTTCTTTCTACCTTTCTTATTTTCAGATCAGCTTTGAGTGAACCTTGACAGAAG
 ATGTTTCGACAGTTTTATCTCTGGACATGTTTAGCTTCAGGGATCATCCTGGGCTCTCTC
 TTTGAAATCTGCTTGGGCCAGTATGATGATGGTAAGGATTGCAAAC TAGCTAGGGGAGGA
 CCACCAGCTACCATAGTTGCTATTGATGAAGAAAGTCGGAATGGTGCAGGTACAATTCTG
 GTGGACAACATGCTGATCAAAGGGACTGCTGGAGGACCAGACCCACCATAAGAATTTCT
 TTAAAGGATAATGTGGATTACTGGGTGTTGATGGATCCTGTTAAGCAAATGCTTTTCCTG
 AACAGCACCGGAAGAGTTCTGGATAGAGATCCACCGATGAACATACACTCCATTGTGGTG
 CAGGTCCAGTGCATCAACAAAAAAGTGGGCACTATTATCTACCATGAAGTGCGAATAGTG
 GTGAGAGACAGGAATGACAACTCACCCACTTTCAAGCATGAAAGCTACTATGCCACAGTG
 AATGAGCTCACTCCAGTTGTTACCACAATATTCACAGGATTTTCAGGAGACAATGGAGCT
 ACAGATATAGATGATGGACCAAATGGACAGATAGAGTATGTTATT CAGTATAATCCAGAT
 GATCCGACATCCAATGACACCTTTGAAATTCCTTAATGTTGACTGGAAATATAGTGTTA
 AGGAAGGCTCAACTATGAAGATAAGACTCGCTACTTTGTCATAATCCAAGCTAATGAC
 CGTGCCCAAAATCTGAATGAGAGGCGAACCACCACCACCTCTCACAGTGGATGTTCTG
 GATGGAGATGACTTGGGTCCAATGTTTCTTCTTGTGTCCTTGTGCCAAACACTCGTGAT
 TGCCGTCCACTCACTTATCAAGCTGCCATACCTGAGTTGAGA ACTCCGGAAGAACTGAAC
 CCCATTATTGTTACGCCACCAATCCAAGCCATTGATCAGGACC GGAATATCAACCGCCA
 TCAGATAGGCCAGGAATCCTCTATTCCATCCTTGTGGTGGGACTCCTGAGGATTACCCA
 CGATTTTTCATATGCATCCTAGGACAGCAGAACTTAGTCTCCTGGAGCCAGTAAACAGA
 GACTTTCACCAAGAAATTTGATTTGGTTATTAAAGGCTGAACAAGACAATGGTCACTCTCT
 CCTGCTTTTGCCGGTCTACACATTGAAATACTGGATGAAAACAATCAAAGTCCATATTTT
 ACAATGCCAGTTATCAAGGCTATATCCTGGAATCTGCCCCAGTGGGAGCAACCATTTCG
 GACAGTCTCAATTTGACCTCACCTTTAAGAATAGTAGCTCTGGACAAGGACATAGAAGAT
 ACAAAGACCCAGAGCTTCACCTTTTTCTGAATGACTACACCTCAGTCTTCACCGTCACA
 CAGACTGGTATTACTCGCTACCTCACCTTACTTCAACCAGTGGACAGGGAAGAACAGCAA
 ACTTACACCTTTTCGATAACAGCATTGTGATGGTGTACAAGAAAGTGAGCCAGTCATCGTC
 AATATTCAAGTGATGGATGCAAATGATAACACGCCAACCTTCCCTGAAATATCCTATGAT
 GTGTATGTTTATACAGACATGAGACCTGGGGACAGTGTATACAGCTCACTGCAGTCGAC
 GCAGACGAAGGGTCAAATGGGGAGATCACATATGAAATCCTTGTGGGGCTCAGGGAGAC
 TTCATCATCAATAAAACAACAGGGCTTATCACCATCGCTCCAGGGGTGGAATGATAGTC
 GGGCGGACTTACGCACTACGGTCCAAGCAGCGGATAATGCTCCTCCTGCAGAGCGAAGG
 AACTCCATCTGCAC TGTGTATATTGAAGTGCTTCCACCAAATAATCAAAGCCCTCCTCGC
 TTCCCACAGCTGATGTATAGCCTTGAAATTAGTGAAGCCATGAGGGTTGGTGCTGTTTTA
 TTAATCTACAGGCAACTGATCGAGAGGAGACTCAATAACATATGCCATTGAGAAATGGA
 GATCCTCAGAGAGTTTAAATCTTTTCAAGAAACACGGGGATTCTAACCTTAGGGAAAGCA
 CTGGACAGGGAAAGCACTGATCGCTACATTCTGATCATCACAGCTTCAGATGGCAGGCCA
 GATGGGACCTCAACTGCCACAGTAAACATAATGGTGACAGATGTCAATGACAATGCTCCA
 GTGTTTGATCCTTATCTGCCAAGAAATTTATCTGTGGTGGAAGAAGAAGCCAATGCCTTT
 GTGGGTCAAGTAAAGCAACAGACCCTGATGCTGGAATAAATGGTCAAGTGCAC TACAGT
 TTGGGTAACTTTAATAATCTTTTCGTATCACATCCAATGGGAGCATTACACAGCAGTG
 AAGCTTAACAGAGAAGTCAGGGACTACTATGAACCTTGTTGTTGTGGCAACAGATGGAGCA
 GTACACCTCGTCATTCAACTCAACCTTGCCATCAAGGTTTGGACATTGATGATAAC
 AGTCTGTGTTTACCAATTCAACATACACTGTCTTGTGTAAGAGAATTTGCCAGCTGGG
 ACTACCATCCTTCAAATAGAGGCCAAAGATGTCGACCTTGGAGCAAATGTGTCTTACCGG
 ATAAGAAGCCCAGAAGTGAAGCACTTTTTTGCCTACATCCATTTACAGGAGA ACTATCG
 CTTTTAAGGAGTTTAGATTATGAGGCATTTCCAGACCAAGAAGCAAGTATCACTTTCTG
 GTAGAGGCCTTTGATATTTATGGAACAATGCCACCTGGTATTGCTACTGTACAGTGATT
 GTAAAGGATATGAATGATTATCCTCCTGTCTTTAGTAAACGAATATACAAAGGGATGGTG
 GCTCCGGATGCAGTCAAGGGTACACCTATCACAAACAGTTTATGCTGAAGATGCAGACCCT
 CCTGGATTACCTGCAAGTCGTGTGAGGTATAGAGTAGATGATGTACAGTTTCTTACCCT
 GCCAGTATTTTTGAAGTGGAAGAAGATTCTGGAAGAGTAATAACACGAGTCAATCTTAAT
 GAAGAACCTACAACAATTTTTAAGTTGGTGGTGGTTGCTTTTGATGATGGGGAGCCTGTG
 ATGTCCAGCAGTGCCACAGTGAAGATTCTTGCTTACATCCTGGTGAGATCCACGCTTC
 ACACAGGAGGAATATAGGCCTCCTCCAGTAAAGTGAACCTGCCACCAAAGGGACCATGGTT
 GGTGTAATTTCTGCTGCTGCCATTAATCAAAGTATTGTGTACTCCATTGTTTCAGGAAAT
 GAAGAAGATACATTTGGAATTAATAACATCACAGGTGTTATCTATGTGAATGGACCTCTG
 GATTATGAGACCAGGACAAGCTATGTACTTCAGAGTCCAAGCTGATTCCTGGAAGTGGTC
 CTTGCCAATCTCCGAGTTCCTTCAAAAAGTAATACAGCTAAAGTATACATTGAGATT CAG
 GATGAAAATAATCATCCCCAGTGTTCAGAAAAAATCTACATCGGAGGTGTATCTGAA
 GATGCAAGAATGTTTACTTCTGTACTCAGAGTGAAGGCTACTGATAAAGATACTGGCAAT

TATAGTGTTCATGGCCTACAGACTCATAATACCACCAATTAAAGAGGGGAAAAGAAGGATTT
GTAGTGGAACATATACAGGGCTTATCAAACTGCTATGCTCTTCCATAATATGAGGAGA
TCCTACTTCAAGTTTCAAGTTATTGCAACTGACGACTATGGGAAGGGACTGAGCGGCAAA
GCCGATGTACTCGTAAGTGTCTCCGTGGTCAATCAGCTGGATATGCAAGTATTGTTTCC
AATGTGCCTCCTACTCTAGTGGAAGGATAGAAAGATCTTACAGAAATCTTGGATCGC
TATGTTTCAGGAACAAATTCCTGGTGCCAAGGTCGTAGTGAGTCCATTGGAGCTCGCCGG
CATGGAGATGCCTTTTCCCTAGAAGATTACACCAATGTGACTTGACTGTCTATGCAATT
GACCCCAAACCAACAGAGCCATCGATAGAAATGAGCTTTTAAAGTTTTTGGATGGCAAA
CTACTTGATATCAATAAAGACTTTTCAGCCGTATTATGGGGAAGGAGGACGCATTCTGGAG
ATCCGGACTCCAGAGGCAGTGACCAGCATTAAAAAGAGAGGAGAAAGTCTAGGATACACA
GAAGGGGCTTGTGGCTCTGGCCTTCATCATCATCCTCTGCTGCATTCTGCCATCTTG
GTGGTTTTGGTCAGCTACAGACAGTTTAAAGTACGTCAAGCTGAGTGACACGTCAAGCT
GAGTGTAACAAGACTGCACGAATTCAGGCCGCATTACCCGCGGCTAAACCAGCAGTGCCG
GCTCCTGCACAGTGGCAGCGCCCCCGCCGCGCCGCTCCGCCAGGTGCGCATCTC
TATGAAGAAGTGGAGACAGCTCAATGTCTTTTCTTTCAAGTCTTTCTCTTCTTACCATT
TTTCAACAAAGCAGGGGAAATAACTCAGTCTCAGAAGACAGGAAACATCAACAAGTTGTG
ATGCCCTTTTCTTCCAATACTATTGAGGCTCACAAGTCAGCTCATGTAGACGGATCACTT
AAGAGCAACAACTGAAGTCTGCAAGAAATTCACATTTCTATCTGATGAGGATGACTTA
AGTGCCCATATCCCTTTTATAAGGAAAACATAAGTCAAGTATCAACAAATTCAGACATT
TCACAGAGAACAGATTTTGTAGACCCATTTTCACCCAAAATACAAGCCAAGAGTAAGTCT
CTGAGGGGCCCCAAGAGAAAAGATTTCAGAGGCTGTGGAGTCAGTCAGTCAGCTTACCCAGG
AGGCTGATGAGGAAAGTTCCAAATAGACCAGAGATCATAGATCTGCAGCAGTGGCAAGGC
ACCAGGCAGAAAGCTGAAAATGAAAACACTGGAATCTGTACAAACAAAAGAGGTAGCAGC
AATCCATTGCTTACAACCTGAAGAGGCAAAATTTGACAGAGAAAGAGGAAATAAGGCAAGGT
GAAACACTGATGATAGAAAGGAACAGAACAGTTGAAATCTCTCTTTCAGACTCTTCATT
TGCTTTCCAGGCCTCACTTCTCATTCTCCACTTTGCCAACTGTTTCAAGAACTGTGGAA
CTCAAATCAGAACCTAATGTCATCAGTTCTCTGCTGAGTGTTTCTTGGAACTTTCTCCT
TCAAGGCCTTGTGTTTTACATTCTTCACTCTCTAGGAGAGAGACACCTATTTGTATGTTA
CCTATTGAAACCGAAAGAAATATTTTTGAAATTTTGGCCATCCACCAAACATCTCTCCT
TCTGCCTGTCCCTTCCCCCTCCTCCTCCTATTTCTCCTCCTTCTCCTCCTCCTGCTCCT
GCTCCTCTTGCTCCTCCTCCTGACATTTCTCCTTTTTCTCTTTTTTGTCTCCTCCTCCTCCT
CCTCCTTCTATCCCTCTTCCCTCTTCCCTCCTCCTACATTTTTTCCACTTTCCGTTTCAACG
TCTGGTCCCCAACACACCTCTTCTACCTCCATTTCCAACCTCCTCTTCTCCTCCTCCTCCT
CCTTCTATTCCCTTGCCCTCCACCTCCTTCAGCTTCATTTCTGTCCACAGAGTGTGTCTGT
ATAACAGGTGTTAAATGCACGACCAACTTGATGCCTGCCGAGAAAATTAAGTCCCTCTATG
ACACAGCTATCAACAACGACAGTGTGTAAAACAGACCCTCAGAGAGAACCAAAAGGCATC
CTCAGACACGTTAAAACTTAGCAGAACTTGAAAAATCAGTAGCTAACATGTACAGTCAA
ATAGAAAAAACTATCTACGCACAAATGTTTCAGAACTTCAAACATGTGCCCCCTCAGAA
GTAACAAATATGGAAATCACATCTGAACAAAACAAGGGGAGTTTGAACAATATTGTGCGAG
GGAAGTGAACAAATCTCACAGTCAATCTACTTCACTGTAATGTTGCTTTTCTTATTTT
AGTCGG

In a search of public sequence databases, the NOV8a nucleic acid sequence, located on chromosome 10 has 557 of 955 bases (58%) identical to a GENBANK-
ID:AF169693|acc:AF169693.1 mRNA from Homo sapiens (Homo sapiens protocadherin 13
5 (PCDH13) mRNA, partial cds). Public nucleotide databases include all GenBank databases
and the GeneSeq patent database.

The disclosed NOV8a polypeptide (SEQ ID NO:18) encoded by SEQ ID NO:17 has 1973 amino acid residues and is presented in Table 8B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV8 has a signal peptide and is likely
10 to be localized extracellularly with a certainty of 0.6760. The most likely cleavage point is between residues 26 and 27.

Table 8B. Encoded NOV8a protein sequence (SEQ ID NO:18).

MFRQFYLTWCLASGIILGSLFEICLGQYDDGKDKCLARGGPPATIVAIDEESRNGAGTIL
VDNMLIKGTAGGPDPTIELSLKDNVDYVWLMDPVKQMLFLNSTGRVLDLDRDPPMNIHSIVV
QVQCINKKVGTTIYHEVRIVVRDRNDNSPTFKHESYYATVNELTPVGTTIFTGFGSDNGA
TDIDGPNQGIEYVIQYNPDPTSNDTFEIPMLTGNIVLRKRLNYEDKTRYFVIIQAND
RAQNLNERRTTTTTLTVDVLDGDDLGPMLPCVLVPNTRDCRPLTYQAAIPELRTPEELN
PIIVTPPIQAIDQDRNIQPPSDRPGILYSILVGGTPEDYPRFFHMHPRTAELSLEPVNR
DFHQKFDLVIKAEQDNGHPLPAFAGLHIEILDENNQSPYFTMPSPYQGYILESAPVGATIS
DSLNLTSPLRIVALDKDIEDTKDELHLFLNDYTSVFTVTQTGITRYLTLLQPVDRREEQQ
TYTFSITAFDGVQSESPVIVNIQVMDANDNTPTFPEISYDVVYVYDMRPGDSVIQLTAVD
ADEGSNGEITYEILVGAQGDPIINKTTGLITIAAGVEMIVGRTYALTQVQADNAPPAERR
NSICTVYIEVLPPNNQSPPRFPQLMYSLEISEAMRVGAVLLNLQATDREGDSITYAIENG
DPQRVFNLSSETTGILTLGKALDRESTDRYILIIITASDGRPDGTSTATVNIMVTDVNDNAP
VFDPYLPRNLSVVEEENAFVGVKATDPDAGINGQVHYSLGNFNNLFRITSNGSIYTAV
KLNREVRDYELVVVATDGAVHPRHSTLTLAIKVLDIDDNSPVFTNSTYTVLVEENLPAG
TTILQIEAKDVDLGANVSYRIRSPVKKHFFALHPFTGELSLRLSDYEAFPDQEASITFL
VEAFDIYGTMPPGIATVTIVIKDMNDYPPVFSKRIYKGMVAPDAVKGTPITTVYAEDADP
PGLPASRVRYRVDVQFPYPASIFEVEEDSGRVITRVNLNEEPTTIFKLVVVAFDDGEPV
MSSSATVKILVLHPGEIPRFTQEEYRPPPVSELATKGTVMGVISAAAINQSIVYSIVSGN
EEDTFGINNITGVIYVNGPLDYETRSTSYVLRVQADSLEVVLANLRVPSKSNATKVYIEIQ
DENNHPPVFQKKFYIGGVSEDARMFTSVLRVKATDKDTGNYSVMAYRLIIPPIKEGKEGF
VVETYTGLIKTAMLFHNMRRSYFKFQVIATDDYGKGLSGKADVLSVSVVNQLDMQVIVS
NVPPTLVEKKIEDLTEILDYVQEIPGAKVVVESIGARRHGDAFSLEDYTKCDLTVYAI
DPQTNRAIDRNELFKFLDGKLLDINKDFQPYGEGGRILEIRTPEAVTSIKKRGESLGYT
EGALLALAFIIILCCIPAILVVLVSYRQFKVRQAECTRQAECKTARIQAALPAKPAVP
APAPVAAPPPPPPPPGAHLYEELGDSSMSFLSSLFLLYHFQQSRGNNSVSEDRKHQQVV
MPFSSNTIEAHKSAHVDGSLKSNKLKSARKFTFLSDEDDLSAHNPLYKENISQVSTNSDI
SQRTDFVDPFSPKIQAKSKSLRGPKEIQRLWSQSVSLPRRLMRKVPNRPEIIDLQQWQG
TRQKAENENTGICTNKRGSNNPLLTTEEANLTEKEEIRQGETLMIEGTEQLKSLSSDSSF
CFPRPHFSFSTLPTVSRTVELKSEPNVISSPAECSLELSPSRPCVLHSSLRRETPICM
PIETERNIFENFAHPPNISPSACPLPPPPPISPSPPPAPAPLAPPPDISPFSFLFCPPPS
PPSIPLPLPPPTFFPLSVSTSGPPTPPLLPFPPTPLPPPPPSIPCPPPPSASFSTECVC
ITGVKCTTNLMPAEKIKSSMTQLSTTTVCKTDPQREPKGILRHVKNLAELEKSVANMYSQ
IEKNYLRTNVSELQTMCPSEVTNMEITSEQNKGSLNNIVEGTEKQSHSQSTSL

A search of sequence databases reveals that the NOV8a amino acid sequence has 1580
of 1846 amino acid residues (85%) identical to, and 1682 of 1846 amino acid residues (91%)
similar to, the 1943 amino acid residue ptnr:TREMBLNEW-ACC:AAG53891 protein from
5 Mus musculus (Mouse) (PROTODADHERIN). Public amino acid databases include the
GenBank databases, SwissProt, PDB and PIR.

NOV8a is expressed in at least brain, lymphoid tissue, placenta. This information was
derived by determining the tissue sources of the sequences that were included in the invention
including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or
10 RACE sources.

NOV8b

A disclosed NOV8b nucleic acid of 6003 nucleotides (also referred to as CG57452-02) encoding a protocadherin-like protein is shown in Table 8C. Putative untranslated regions upstream and/or downstream from the coding region, if any, are underlined, and the start and stop codons are in bold letters.

Table 8C. NOV8b nucleotide sequence (SEQ ID NO:19).

CAAATGTTTCTTTCTACCTTTTCTTATTTTCAGATCAGCTTTGAGTGAACCTTTGACAGAAG
ATGTTTTCGACAGTTTATCTCTGGACATGTTTAGCTTCAGGGATCATCCTGGGCTCTCTC
 TTTGAAATCTGCTTGGGCCAGTATGATGATGGTAAGGATTGCAAACCTAGCTAGGGGAGGA
 CCACCAGCTACCATAGTTGCTATTGATGAAGAAAGTCGGAATGGTGCAGGTACAATTCTG
 GTGGACAACATGCTGATCAAAGGGACTGCTGGAGGACCAGACCCACCATAGAACCTTTCT
 TTAAAGGATAATGTGGATTACTGGGTGTTGATGGATCCTGTTAAGCAAATGCTTTTCCTG
 AACAGCACCGGAAGAGTTCTGGATAGAGATCCACCGATGAACATACACTCCATTGTGGTG
 CAGGTCCAGTGCATCAACAAAAAGTGGGCACTATTATCTACCATGAAGTGCGAATAGTG
 GTGAGAGACAGGAATGACAACCTCACCCACTTTCAAGCATGAAAGCTACTATGCCACAGTG
 AATGAGCTCACTCCAGTTGGTACCACAATATTCACAGGATTTTCAGGAGACAAATGGAGCT
 ACAGATATAGATGATGGACCAAAATGGACAGATAGAGTATGTTATTTCAGTATAATCCAGAT
 GATCCGACATCCAATGACACCTTTGAAATTCCTTAATGTTGACTGGAAATATAGTGTTA
 AGGAAGAGGCTCAACTATGAAGATAAGACTCGCTACTTTGTCATAATCCAAGCTAATGAC
 CGTGCCCAAAATCTGAATGAGAGGCGAACCACCACCACCTCTCACAGTGGATGTTCTG
 GATGGAGATGACTTGGGTCCAATGTTTCTTCTTGTGTCCTTGTGCCAAACACTCGTGAT
 TGCCGTCCACTCACTTATCAAGCTGCCATACCTGAGTTGAGAACTCCGGAAGAACTGAAC
 CCCATTATTGTTACGCCACCAATCCAAGCCATTGATCAGGACCGGAATATTCAACGCCA
 TCAGATAGGCCAGGAATCCTCTATTCCATCCTTGTGGGACTCCTGAGGATTACCCACGA
 TTTTTCATATGCATCCTAGGACAGCAGAACTTAGTCTCCTGGAGCCAGTAAACAGAGAC
 TTTTACCAGAAATTTGATTTGGTTATTAAGGCTGAACAAGACAATGGTCATCCTCTTCT
 GCCTTTGCCAGTCTACACATTGAAATACTGGATGAAACAATCAAAGTCCATATTTTACA
 ATGCCCAGTTATCAAGGCTATATCCTGGAATCTGCCCCAGTGGGAGCAACCATTTCGGAC
 AGTCTCAATTTGACTTCACCTTTAAGAATAGTAGCTCTGGACAAGGACATAGAAGATACA
 AAAGACCCAGAGCTTCACCTTTTCTGAATGACTACACCTCAGTCTTCACCGTCACACAG
 ACTGGTATTACTCGCTACCTCAGCTTACTTCAACCAGTGGACAGGGAAGAAGACAGCAAAC
 TACACCTTTTTCGATAACAGCATTGATGGTGTACAAGAAAGTGAGCCAGTCATCGTCAAT
 ATTCAAGTGATGGATGCAAATGATAACACGCCAACCTTCCCTGAAATATCCTATGATGTG
 TATGTTTATACAGACATGAGACCTGGGGACAGTGTACATACAGCTCACTGCAGTCGACGCA
 GACGAAGGGTCAAATGGGGAGATCACATATGAAATCCTTGTGGGGCTCAGGGAGACTTC
 ATCATCAATAAAACAACAGGGCTTATCACCATCGCTCCAGGGGTGGAATGATAGTCGGG
 CGGACTTACGCACTCACGGTCCAAGCAGCGGATAATGCTCCTCCTGCAGAGCGAAGGAAC
 TCCATCTGCACTGTGTATATTGAAGTGCTTCCACCAAATAATCAAAGCCCTCCTCGCTTC
 CCACAGCTGATGTATAGCCTTGAAATTAGTGAAGCCATGAGGGTTGGTGCTGTTTTATTA
 AATCTACAGGCAACTGATCGAGAGGGAGACTCAATAACATATGCCATTGAGAATGGAGAT
 CCTCAGAGAGTTTTTAATCTTTTCAGAAACCACGGGGATTCTAACCTTAGGGAAAGCACTG
 GACAGGGAAAGCACTGATCGCTACATTCTGATCATCACAGCTTCAGATGGCAGGCCAGAT
 GGGACCTCAACTGCCACAGTAAACATAGTGGTGACAGATGTCAATGACAATGCTCCAGTG
 TTTGATCCTTATCTGCCAAGAAATTTATCTGTGGTGGAAGAAGAAGCCAATGCCTTTGTG
 GGTCAAGTAAAGCAACAGACCTTGATGCTGGAATAAATGGTCAAGTGCACCTACAGTTTG
 GGTAACTTTAATAATCTTTTTCGTATCACATCCAATGGGAGCAATTTACACAGCTGAAG
 CTTAACAGAGAAGTCAGGGACTACTATGAACCTTGTGTTGGCAACAGATGGAGCAGTA
 CACCTCTGTCATTCAACTCTAACCTTGGCCATCAAGGTTTTGGACATTGATGATAACAGT
 CCTGTGTTACCAATTCAACATACACTGTCCTTGTGTAAGAGAATTTGCCAGCTGGGACT
 ACCATCCTTCAAATAGAGGCCAAAGATGTCGACCTTGGAGCAAATGTGTCTTACCGGATA
 AGAAGCCCAGAAAGTGAAGCACTTTTTTGCACTACATCCATTTACAGGAGAACTATCGCTT
 TTAAGGAGTTTAGATTATGAGGCATTTCCAGACCAAGAAGCAAGTATCACTTTTCTGGTA
 GAGGCTTTTGATATTTATGGAACAATGCCACCTGGTATTGCTACTGTACAGTGATTGTA
 AAGGATATGAATGATTATCCTCCTGCTTTTAGTAAACGAATATACAAAGGGATGGTGGCT
 CCGGATGCAGTCAAGGGTACACCTATCACAACAGTTTATGCTGAAGATGCAGACCTCCT

CGG

5

The disclosed NOV8b polypeptide (SEQ ID NO:20) encoded by SEQ ID NO:19 has 1972 amino acid residues and is presented in Table 8D using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV8b has a signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.6760. The most likely cleavage site is between amino acids 26 and 27.

Table 8D. Encoded NOV8b protein sequence (SEQ ID NO:20).
MFRQFYLTCLASGIILGSLFEICLGQYDDGKCKLARGGPPATIVAIDEESRNGAGTIL VDNMLIKGTAGGPDPTIELSLKDNVDYVWLMDPVKQMLFLNSTGRVLDLDRDPPMNIHSIVV QVQCINKKVGTTIYHEVRIVVRDRNDNSPTFKHESYYATVNELTPVGTTIFTGFGSDNGA TDIDDGPNQGIEYVIQYNPDPTSNDTFEIPLMLTGNIVLRKRLNYEDKTRYFVIIQAND RAQNLNERRTTTTTLTVDVLDGDDLGPMFLPCVLVPNTRDCRPLTYQAAIPELRTPEELN PIIVTPPIQAIDQDRNIQPPSDRPGILYSILVGTPEDYPRFFHMHPRTAELSLLEPVNRD FHQKFDLVIKAEQDNGHPLPAFASLHIEILDENNQSPYFTMPSYQGYILESAPVGATISD SLNLTSPLRIVALDKDIEDTKDPELHLFLNDYTSVFTVTQTGITRYLSLLQPVDRREEQQT YTFISITAFDGVQSEFVIVNIQVMDANDNTPTFPEISYDVVYVYDMRPGDSVIQLTAVDA DEGSNGEITYEILVGAQGDFFINKTTGLITIAAGVEMIVGRTYALTQQAADNAPPARRN SICTVYIEVLPPNNQSPPRFPQMLYSLEISEAMRVGAVLLNLQATDREGDSITYAIENGD PQRVFNLSETTGILTLGKALDRESTDRYILIIITASDGRPDGTSTATVNIIVTVDVNDNAPV FDPYLPRLNLSVVEEEANAFVGVQVKATDPDAGINGQVHYSLGNFNNLFRITSNGSIYTAVK LNREVRDYYELVVVATDGAHVPRHSTLTLLAIKVLDDIDNSPVFTNSTYTVLVEENLPAGT TILQIEAKDVDLGANVSYRIRSPEVKHFFALHPFTGELSLRLSLDYEAFFPDQEASITFLV EAFDIYGTMPPGIATVTVIVKDMNDYPPVFSKRIYKGMVAPDAVKGTPITTVYAEDADPP GLPASRVRYRVDDVQFFYPASIFEVEEDSGRVTIRVNLNEEPTTIFKLVVVAFDDGEPVM SSSATKILVLHPGEIPRFTQEEYRPPPVSELATKGMTVGVISAAAINQSIYVSIVSGNE EDTFGINNITGVIYVNGPLDYETRSTYVLRVQADSLEVVLANLRVPSKSNKAKVYIEIQD ENNHPVFQKKFYIGGVSEDAARMFTSVLRVKATDKDTGNYSVMAYRLIIPPIKEGKEGFV VETYTGLIKLTAMLFHNMRRSYFKFQVIATDDYGKGLSGKADVLVSVSVVNQLDMQVIVSN VPPTLVEKKIEDLTEILDYVQEQIPGAKVVVESIGARRHGDAFSLDYTKCDLTVYAIID PQTNRAIDRNELEFKFLDGKLLDINKDFQPYGEGGRILEIRTPEAVTSIKKRGESLGYTE GALLALAFIIILCCIPAILVVLVSYRQFKVRQAECTRQAECTKTARIQAALPAKPAVPA PAPVAAPPPPPPPPPGAHLYEELGDSSMSFLSSLFLLYHFQQSRGNNSVSEDKRHHQVVM PFSSNTIEAHKSAHVDGSLKSNKLKSARKFTFLSDEDDLSAHNPLYKENISQVSTNSDIS QRTDFVDPFSPKIQAKSKSLRGPKEIKRLWSQSVSLPRRLMRKVPNRPEIIDLQQWQGT RQKAENENTGICTNKRGSNPLLTTEEANLTEKEEIRQETLMIEGTEQLKSLSSDSSFC FPRPHFSFSTLPTVSRTVELKSEPNVISSPAECSLELSPSRPCVLHSSLSRRETPICMLP IETERNIFENFAHPPNISPSACPLPPPPPISPSPPPAPAPLAPPPDISPFLFCPPPPSP PSIPLPLPPPTFFPLSVSTSGPPTPPLLPPFPTPLPPPPPSIPCPPPPSASFSTECVCI TGVKCTTNLMPAEKIKSMTQLSTTTVCKTDPQREPKGILRHVKNLAELEKSANMYSQI EKNYLRTNVSELQTMCPSEVTNMEITSEQNKGLNNIVEGTEKQSHSQSTSL

A search of sequence databases reveals that the NOV8b amino acid sequence has 3708 of 4369 bases (84%) identical to a gb:GENBANK-ID:AF281899|acc:AF281899.1 mRNA from Mus musculus (Mus musculus protocadherin (av) mRNA, complete cds). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV8b is expressed in at least adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine,

spinal cord, spleen, stomach, testis, thyroid, trachea and uterus. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

- 5 The disclosed NOV8a polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 8E.

Table 8E. BLAST results for NOV8a					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
<u>gi 16933555 ref NP_149045.2 </u> (NM_033056)	protocadherin 15 precursor; Usher syndrome 1F (autosomal recessive, severe) [Homo sapiens]	1955	1954/1973 (99%)	1955/1973 (99%)	0.0
<u>gi 14581464 gb AAK31581.1 </u> (AY029237)	protocadherin 15 [Homo sapiens]	1955	1955/1973 (98%)	1955/1973 (98%)	0.0
<u>gi 15072441 gb AAK31581.1 </u> (AY029205)	protocadherin 15 [Homo sapiens]	1955	1949/1973 (98%)	1955/1973 (98%)	0.0
<u>gi 12963485 ref NP_075604.1 </u> (NM_023115)	protocadherin 15; Ames waltzer [Mus musculus]	1943	1626/1977 (82%), Positives = 1745/1977 (88%)	1626/1977 (82%), Positives = 1745/1977 (88%)	0.0
<u>gi 18574084 ref XP_053625.2 </u> (XM_053625)	protocadherin 15 precursor [Homo sapiens]	936	924/928 (99%)	925/928 (99%)	0.0

- 10 Table 8F lists the domain descriptions from DOMAIN analysis results against NOV8. This indicates that the NOV8 sequence has properties similar to those of other proteins known to contain this domain.

Table 8F. Domain Analysis of NOV8

gnl|Smart|smart00112, CA, Cadherin repeats.; Cadherins are glycoproteins involved in Ca²⁺-mediated cell-cell adhesion. Cadherin domains occur as repeats in the extracellular regions which are thought to mediate cell-cell contact when bound to calcium.

CD-Length = 82 residues, 100.0% aligned

Score = 83.2 bits (204), Expect = 1e-16

The disclosed NOV8 polypeptides are members of the protocadherin family, which in turn is one of the six subfamilies of the cadherin superfamily. Cadherins are membrane-associated glycoproteins that mediate cell-cell interactions in a calcium-dependent fashion.

5 Protocadherins may act as cell-cell recognition molecules and may be involved in signal transduction cascades.

The disclosed NOV8 polypeptides have homology to the mouse protocadherin whose mutant version causes the Ames waltzer mouse phenotype, which includes deafness and a balance disorder due to degeneration of the neuroepithelium of the inner ear. Mutant mice
10 show abnormal stereocilia in the inner ear at a very early age. The gene of invention may therefore have a role in developmental processes, cellular communication and disease processes such as cancer.

The disclosed NOV8 nucleic acids of the invention encode a protocadherin-like protein includes the nucleic acid whose sequence is provided in Table 8A or 8C or a fragment
15 thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 8A or 8C while still encoding a protein that maintains its protocadherin-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are
20 complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical
25 stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 42 percent of the bases may be so changed.

The disclosed NOV8 protein of the invention includes the protocadherin-like protein whose sequence is provided in Table 8B or 8D. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 8B or 8D while still encoding a protein that maintains its protocadherin-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 15 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this protocadherin-like protein (NOV8) may function as a member of a "protocadherin family". Therefore, the NOV8 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV8 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the protocadherin-like protein (NOV8) may be useful in gene therapy, and the protocadherin-like protein (NOV8) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neurodegeneration hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, graft versus host disease (GVHD), lymphoedema, hearing loss, tinnitus, balance disorders, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, transplantation, cancer, tissue degeneration, bacterial/viral/parasitic infection, or other pathologies or conditions. The NOV8 nucleic acid

encoding the protocadherin-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV8 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV8 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV8 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV9

A disclosed NOV9 nucleic acid of 13700 nucleotides (also referred to as CG57625-01) encoding a protocadherin-like protein is shown in Table 9A. Putative untranslated regions upstream and/or downstream from the coding region, if any, are underlined, and the start and stop codons are in bold letters.

Table 9A. NOV9 nucleotide sequence (SEQ ID NO:21).

TCTCTCTTTCTTCCCTCCAGGATGGAAGTATGATGTGATGGATATAATTATGGGACACTG
TGTGGGCACACGGCTCCTGCTTGTGGCCTCATCCTCCTGCTTTTCAAGCTTTTGGCCAC
TGTCTCCCAGGGGCTGCCAGGGACTGGACCCCTGGGCTTCCACTTCACACATTCCTATTA
TAATGCTACCGTGTATGAGAACTCAGCAGCAAGGACCTACGTCAACAGCCAGAGTAGAAT
GGGCATCACCTTAATAGATCTATCCTGGGATATCAAATACAGAATAGTGTCCGGAGACGA
GGAAGGCTTTTCAAAGCAGAGGAAGTCATCATTGCAGATTCTGTCTTCTCAGAATAAG
AACTAAAGGTGGCAATTCTGCCATATTAAATAGGGAAATCCAGGATAATTATTTATTGAT
AGTAAAGGTTCTGTCAGAGGAGAGGATTTGGAAGCATGGACCAAAGTGAATATACAGGT
TTTAGATATGAATGATCTGAGACCTTTGTTTTTCAACCAACATACTCTGTTACCATAGC
AGAAAGCACACCTCTAAGGACTAGTGTGCCCCAGGTGACTGCAACAGACGCAGATATTGG
TTCCAATGGAGAATTCTACTACTACTTTAAAAATAAAGTTGATCTCTTTTCAGTTCACCC
CACGAGTGGTGTCTCTCCTTAAGTGGTCGATTAAATTATGATGAAAAGAATAGGTATGA
TCTGGAATTTTGGCTGTGGACCGGGGAATGAACTGTATGGGAACAATGGAGTGAGCAG
TACTGCAAAGCTTTATGTTTACATTGAGCGCATAAATGAACATGCCCCAACAAATCCATGT
AGTCACTCATGTTCTTTCTCGTTGGAAAAAGAGCCAACATATGCAGTGGTGACAGTTGA
TGACTTAGATGATGGAGCGAATGGAGAGATCGAATCTGTTTCCATTGTGGCTGGGGATCC
TTTAGATCAGTTCTTCTGGCTAAGGAAGGAAAGTGGTTGAATGAGTACAAGATTAAGGA
GAGGAAGCAGATTGACTGGGAGAGCTTTCCCTATGGCTACAATCTCACTCTTCAAGCAAA
AGACAAGGATCTCCTCAAAAATGTTTCAAGCATTAAAGGCAGTCTACATTGGCAACCCAC
AAGAGACACTGTCCCCATTAGATTTGAAAAAGAAGTGTACGATGTGAGCATAAGTGAATT
TTCCCTCTCGTGTGCTATAGTAAATTAAGTCTGAACCGATAGATGTGGA
ATACAAATTATCTCCTGGTGAGGATGCAGTGTACTTTAAATTAATCCTCGGTGCGGTCT
GATTGTTACAGCACGGCCACTGAATACTGTTAAGAAGGAGGTTTATAAACTGGAGGTGAC
AAACAAGGAAGGAGATTTAAAGCACAGGTCACCATCAGCATAGAAGATGCAATGACCA
CACCCAGAAATTCAGCAACCACTGTATGATGCTTATGTGAATGAAAGTGTCCAGTGGG
AACCAGCGTTCTAACAGTTTTCAGCTTCTGATAAGGATAAAGGAGAAAATGGGTACATCAC
CTATAGTATCGCTAGCCTGAATTTGTTACCATTTGTCATTAATCAGTTTACAGGTGTTAT

TAGCACAACTGAAGAACTGGATTTTGAATCCTCCCCAGAAATTTACAGATTCATTGTTAG
AGCCTCTGACTGGGGTTCACCATACCGCCATGAAAGTGAGGTCAATGTGACTATTTCGAAT
AGGAAATGTCAACGACAACAGCCCTCTCTTTGAAAAAGTGGCTTGCCAGGGAGTTATTTTC
ATATGACTTTCCAGTTGGTGGTCACATCACAGCAGTCTCAGCGATCGATATCGATGAACT
TGAAGTTGTAAAGTACAAAATCATTTCTGGAAATGAACTTGGCTTCTTTTATTTAAACCC
AGATTCTGGTGTTTTACAGCTTAAAAAATCACTGACAAATCTGGCATTAAAAATGGCAA
TTTTGCCCTCAGAAATTACAGCAACTGATGGAGAGAATCTTGCAGACCCCATGTCTATTAA
CATTTTCAGTCTTACATGGGAAAGTGTCTTCAAAGAGCTTCAGTTGCAGAGAAACTCGTGT
GGCTCAAAAGCTGGCAGAGAACTACTCATTAAAGGCAAAAGCAAATGGGAAACTGAATCT
GGAAGATGGATTTCTTGACTTTTATTCAATTAATAGACAGGGACCATATTTTGACAAGTC
TTTTCTCTTGATGTGGCTGTAAAGGAGGATCTGCCAGTTGGTGCTAACATTCTGAAGAT
TAAAGCCTATGATGCCGACTCTGGCTTCAATGGAAAAGTGCTATTTACAATATCAGATGG
AAATACGGATAGTTGCTTTAATATTGATATGGAGACTGGGCAGCTTAAAGTCCTTATGCC
CATGGATCGAGAACACACAGACCTCTATCTCCTTAATATCACCATCTATGACTTAGGTAA
TCCACAGAAATCGTCATGGAGACTGCTGACCATCAATGTGGAGGATGCTAATGACAATAG
CCCAGTTTTTTATTCAAGACAGTTACTCAGTTAACATTCTTGAAAGTTTCAGGCATTGGTAC
TGAAATCATTCAAGTGAAGCCAGAGACAAAGACTTAGGTTCTAATGGTGAAGTGACTTA
CTCAGTCTTGACAGATACACAGCAGTTTGGCCATCAATAGCTCAACTGGAATCGTTTATGT
AGCCGACCAGTTGGACCGGGAATCCAAAGCCAATTATTCTTTGAAAAATAGAAGCCAGGGA
CAAGGCAGAGAGTGGTCAGCAGCTGTTTTTCAAGTTGTCACTCTTAAAGTTTTTTTAGATGA
TGTCATGACTGCTCCCCAGCTTTTCAATCCAGTAGCTATAGTGTGAAGGTTCTTGAAGA
TCTCCCTGTTGGCAGTGTCTATTGCTTGGCTTGAGACCCATGATCCAGATCTTGGACTGGG
GGGTCAAGTGCGCTATTCTTTGGTCAATGACTATAATGGGAGATTTGAAATAGATAAAGC
AAGTGGTGCCATCCGCTTGAGCAAAGAGCTTGATTATGAGAAACAGCAGTTCTATAACCT
TACTGTGCGGGCCAAAGACAAAGGGCGGCTGTCTCTCTGTCATCTGTTTCTTGTGA
GGTGAAGTGGTGGATGTCAATGAAAACCTCCACACTCCCTATTTCCAGACTTTGCTGT
TGTTGGATCTGTAAAGGAAAACTCACGCATTGGAACAAGCGTGCTGCAGGTGACTGCTCG
AGATGAAGACTCCGGAAGGGATGGAGAGATCCAGTACTCCATCAGGGATGGCAGTGGTCT
TGGAAGGTTTCAGTATAGACGACGAGAGTGGGGTCATCACTGCCGCAGACATTCTTGATCG
GGAGACAATGGGGTCATACTGGCTAACAGTGTATGCCACAGACAGGGGCGTTGTTCCACT
CTACTCCACCATTGAGGTCTACATTGAAGTTGAAGATGTGAATGACAATGCCCCGCTGAC
CTCAGAACCCCTATATATTATCTGGTCAATGGATAAACATCCGAAGGACGTATCTGTCTAT
TCAGATCCGGCTGAAGATCCTGACTCCAGTTCCAATGAAAACTGACATACAGGATTAC
AAGTGGAAATCCTCAGAAATTTTTTGTGCATCAATATCAAAACAGGGCTGATTACAACAAC
TTCAAGGAAATTTGGATCGAGAACAGCAGGCAGAACATTTTCTGGAGGTGACTGTGACAGA
TGGTGGTCCCTCTCCAAAACAGTCAACCATTGGGGTGGTGGTTCAGGTTCTAGATGAAAA
TGACAACAAGCCCCAGTTCCAGAGAAAGGTCTACCAGATCAAGCTGCCAGAACGTGACCG
AAAGAAGAGAGGAGAACCGATTTACAGGGCTTTTGCATTTGATAGAGATGAGGGCCCCAA
CGCAGAAATCTCTACAGTATTGTGGATGGGAATGATGACGGAAGTTCTTTATTGACCC
TAAAGTCCGGATGGTTTCTTCTAGAAAGCAGTTTACAGCAGGCAGTTATGACATCCCTAAC
GATAAAGGCAGTGGACAATGGGCGCCACAGAAATCCTCCACGGCCCCGCTCCACATTGA
ATGGATTAAGAAACCACCCCTTACCTATACCATTTGACCTTCGATGAGCCGTTTTTATAA
CTTCACAGTCATGGAAGTGATAGAGTGACTGAAATTGTAGGGGTGGTGTCTGTGCAGCC
AGCTAACACCCCTCTGTGGTTTGACATAGTTGGGGGGAATTTTGACAGCGCTTTTGATGC
AGAGAAGGGTGTGGGACAATTGTCTATCGCAAAACCTTTGGATGCAGAGCAGAGGTCCAT
CTATAATATGAGTGTGGAAGTCACCGATGGGACAAATGTTGCTGTTACTCAGGTATTTAT
CAAAGTGTGGATAATAATGATAATGGCCCAGAATTCTCTCAGCCGAATTACGATGTGAC
AATTTCCGAGGATGTGCTTCCAGACACGGAGATCCTGCAGATTGAAGCCACAGATAGAGA
TGAGAAGCACAGCTGAGCTACACTGTTTCATAGCAGCATCGACTCCATCAGCATGAGAAA
ATTCCGGATTGACCCTAGCACTGGCGTGCTCTATACTGCCGAGAGGCTGGACCATGAGGC
CCAGGACAAGCACATTCTCAACATAATGGTCAGAGATCAGGAGTTTCCTTATCGAAGAAA
CTTGGCCCCGAGTCATTGTGAATGTGGAGGATGCTAATGATCACAGTCCTTATTTTACCAA
CCCAGTGTATGAAGCGTCTGTGTTTGAATCTGCTGCTCTGGGATCAGCTGTTCTGCAAGT
GACGGCTCTGGACAAAAGACAAAGGAGAAAATGCAGAACTCATATATACCATAGAAGCAGG
GAACACTGGGAACATGTTAAGATCGAACCGGTCTTAGGCATCATCACCATTGCAAGA
ACCAGACATGACGACGATGGGTGAGTTTGTCTTATCCATCAAAGTCACAGATCAGGGATC
CCCGCCAATGTCTGCTACTGCAATTGTGCGCATTTCCGTCACCATGTCTGACAATTCTCA
CCCCAAGTTCAATCACAAGACTACCAAGCAGAAGTAAATGAAAATGTTGACATTGGAAC
ATCAGTCATTCTAATCTTGCCATCAGTCAATCTACCCCTATTATGAAGTCAAAGATGG
AGACATTAATGGGATCTTTACCATAAATCCATATTCTGGAGTCATCACCCTCAGAAGGC
CCTGGATTATGAGCGCACATCTCTTATCAACTCATCATTAGGCCACCAATATGGCAGG
AATGGCTTCCAATGCTACAGTCAATATTAGATTGTTGATGAAAATGATAATGCCCCAGT

TTTTTCTCTTTTCTCTCAATACTCAGGCAGCCCTAAGTGAGGCTGCCCAAATTAATAGCATTGT
CAGGAGCTTGGATAACAGCCCACTGGTGATTTCGAGCCACAGATGCTGACAGCAACCGGAA
TGCTCTGCTTGTGTATCAGATTGTGGAGTCAACAGCAAAAAAGTTTTTTACGGTGGACTC
CAGTACAGGTGCAATCAGAACAAATGCCAACCTGGACCATTGAAACCATTTGCCATTTCCA
TTTTCTAGTGCATGTGAGAGACAGTGGTAGCCCCCACTGACTGCTGACAGAGATCCCCTTGA
AGTCAACATTTGAGGTGACAGATGTGAATGATAACCCACCTGTTTTTACTCAGGCTGTGTT
TGAGACTATCTTACTTCTACCTACCTATGTTGGAGTGAGGTTCTGAAAGTTAGTGCCAC
AGATCCTGACTCTGAGGTACCCCTGAACTGACATACAGCCTAATGGAAGGCAGTTTGGAA
TCATTTTTTAATTGACTCAAACAGTGGAGTACTTACCATAAAAAACAACAACCTCTCCAA
GGATCACTACATGCTGATAGTTAAGGTGTCTGATGGAAAAGTTCTACAGTACCTCCATGGT
CACCATCACTGGTTAAAGAAGCCATTGGACAGCGGCCCTCCACTTTACACAAAGCTCTTATT
CACCTCAATCTCAGAGAACACACTAACATAACCAAAGTTGCTATTGTCAATGCAGTTTGG
AAATCGCCTTAATGAGCCCTTAAATACAGCATCTTAAACCCAGGAAATAAGTTCAAGAT
AAAATCTACCTCAGGGGTCAATCAGACGACTGGAGTCCCCTTTGACCGTGAAGAACAAGA
GTTATATGAGCTGGTGGTAGAAGCCAGCCGTGAGCTGGACCATCTGCGTGTGGCCAGAGT
GGTGGTCAGGGTTAACATTGAAGACATAAATGACAATTCTCCAGTCTTTGTGGGCCTCCC
ATACTATGCTGCTGTTCAAGTGGATGCGGAACCCGGGACTCTGATTTATCAGGTGACAGC
CATTGACAAAAGATAAAAGGTCCAAATGGAGAAGTGACCTATGTCTCTGAGGATGACTATGG
CCACTTTGAAATTAACCCATAATTGAGGGAATGTTATTTTAAAGGAAGCATTCAACTTGA
CTTGTGCCAATCATTGAGTATGAGTCACTCACCATCCTAGCCAAGGATGGCGGAAACCTTCTTT
GTCTACATCTGTGGAGCTTCCCATCACTATTGTCAACAAAGCAATGCCTGTGTTTGATAA
GCCCTTTTATACAGCATCTGTCAATGAAGACATCAGAATGAACACACCCATCCTAAGCAT
CAATGCCACCAGTCCAGAAGGCCAAGGCATCATATATCATTATCGATGGGGACCCCTTT
TAAACAGTTTAAACATTGACTTTGACACTGGGGTCTGAAAGTTGTTAGCCCTTTGGATTA
TGAAGTTACATCTGCTTACAAGCTGACAATAAGAGCCAGCGACGCCCTTACTGGTGCTAG
GGCTGAAGTCACTGTGACTTGCTAGTTTAAATGATGTAATGACAACCCCTTATTTTCGA
TCAGCCTACATACAATAACAACACTATCAGAAGCATCTCTTATTTGGGACACCTGTTTTTACA
AGTTGTCTCTATTGATGCAGACTCAGAAAACAATAAAATGGTACATTATCAGATTGTCCA
GGATACCTACAATAGCACAGATTATTTTTCATAGATAGCTCAAGTGGCTTAATCCTGAC
AGCAGCAATGCTGGACCATGAGTTAGTACAACACTGCACTTTGAAAGTCAGATCAATAGA
TAGTGGCTTCCCATCACTGAGCAGTGAGGTTCTCGTTTCATATCTACATCTCTGATGTA
TGACAAACCCCTCCAGTTTTTTAATCAGCTCATTTATGAGTCAATATGAGTGAATTAGCCCC
CCGGGACCCATTTGTAACTCTGTGTACAAGCCTCTGATGCAGACAGCTCTGATTTTGGCCG
GTTTGGAAATATAGCATTTTATCTGGGAATGACCGGACGAGCTTTCTGATGGACAGCAAGAG
TGGAGTTATCACATTGTCCAACCATCGGAAGCAGCGGATGGAGCCTCTGTACAGTCTCAA
TGTGTCTGTCTCTGATGGGTTGTTTACCAGCACTGCACAGGTGCATATTAGGGTACTTGG
GGCTAACTTGTACAGCCCTGCCTTTTTCACAAAGCACATACGTAGCTGAGGTGAGAGAGAA
CGTGGCTGCAGGAACAAAGGTAATTATGTTTCGAGCCACAGATGGTGTACAGGGACTTA
TGCGCAGATCAGCTATGCCATCATCAATGACTTTGGCAAGGATCGATTCCTCATAGACAG
CAATGGGCGAGTGCATCACACAGAAAGGCTAGACCGGGAACCCCTAGAAAGGGGATGT
TAGATTTTTTGTGAGGGCCCTTGATGGTGGAGGGAGAAACAACTTTCTGCACTGTGAGAGT
GATTGTTGTGGATGAAATGACAATGCTCCCCAGTTCATGACAGTGGAATATAGAGCCAG
TGTGAGGGCAGATGTTGGAAGGGGCCACTTGGTCACTCAAGTTCAAGCCATAGATCCCGA
TGATGGAGCAAATTCAGGATTACTTATCCCTCTATAGCGAGGCCTCTGTTTCAGTGGC
CGACCTCCTGGAAATCGATCCTGACAAATGGCTGGATGGTCACAAAGGGTAATTTTAACCA
GCTGAAAAATACAGTGCCTTTGCTTTTGTCAAAGCAGTAGATGGGGCATCCCAGTAAA
GCACCTCCTCATTCCTGTCTATATCCACGTCTTGCCCTGAAACGTTCTTGCCATCACT
CACCAGTCTCAGTATTCTTTTACCATTGCAGAAGATACAGCCATTGGGAGTACAGTGGAA
CACCTTGAGGATTTTGCCAGTCAAGATGTCTGGTTCAGCACAGTTAATGGGGAACGGCC
AGAAAATAACAAAGGGGGCATATTTCGTATAGAACAGGAAACAGGCACATTAAGCTTGA
CAAACGCCTTGACCGTGAAACCAGCCAGCTTTTCCACTTTAAAGTAGCAGCCACTATACC
CCTGGACAAAGTAGACATTGTGTTTACTGTGGATGTAGATATCAAGGTATTGGATTGAA
TGACAAACAGCCAGTCTTTGAACTTCAAGCTATGACACCATTATAATGGAAGGGATGCC
TGTTGGCAGCAAACTCACACAAGTGAGAGCTATTGATATGGACTGGGGAGCCAAATGGACA
AGTCACTTACTCCCTCCACTCGGATTTCCAGCCGAAAGGTAATGGAAGCATTCAATAT
TGACAGCAACACGGGCTGGATCAGTACCTTGAAGGACCTAGATCACGAGACAGACCCAC
ATTACCTTCTCTGTGGTGGCCTCTGACCTTGGAGAGGCATTCTCTCTTTCCTCCACGGC
CTTGGTCTCTGTGAGAGTGACAGATATAAATGACAATGCACCAGTCTTCGCGCAGGAAGT
GTACCGAGGGAAATGTGAAGGAGAGCGACCCACCGGGCAGGTTGGTAGCCGTCTTCAGCAC
CTGGGACAGAGACATCCGACGTTTAACTGCCAAGTGAGCTACCATATTACAGTTGGAAAA
CCCTCGAGGAAGGTTTGTCTTGGCCCTGGTGCAAAGTGAGTGGAAGGTTCTATGTAAGAG
GCCTCTAGACAGAGAAGAACAGGACATTTACTTTTCTCAATATCACTGCCACTGATGGGCT

TTTTGTCACACAGGCCATGGTGAAGTGAGCGTCAGTGATGTGAATGACAATAGCCCAGT
GTGTGATCAGGTTGCATATACAGCATTACTTCCTGAAGACATTCCATCAAATAAAATCAT
CCTGAAAGTCAGTGCAAAGGATGCTGATATTGGATCCAATGGATATATACGATACTCACT
CTATGGATCTGGAAACAGTGAATTTTTCTAGATCCAGAAAGTGGCGAGTTAAAAACCTT
GGCTCTGTTGGACCGGGAGAGGATCCCCGTGTACAGCCTGATGGCCAAGGCCACTGACGG
GGTGGCAGGTTCTGCCAGTCCAACATCCACCTAATCCTGGAGGATGTGAATGATAACCC
CCCTGTGTTTTCTTCTGACCACTACAACACCTGTGTCTATGAGAACACAGCCACCAAGGC
TCTGTGACCAGAGTTCAAGCCGTGGACCCCGACATTGGCATCAATAGGAAGGTCGTGTA
CTCCCTGGCAGACTCAGCTGGTGGGGTCTTCTCCATTGACAGCTCATCTGGCATCATCAT
CCTGGAGCAGCCACTGGACCGTGAGCAGCAGTCTTCGTACAACATCAGCGTGCGGGCCAC
TGACCAGAGTCCTGGACAGTCCCTGTCTCTCTCACTACTGTCAACATCACCGTCTCTGGA
CATTAATGACAACCCCTGTGTTTGGAGGAGGGACTACCTGGTGACGGTGCCTGAGGA
CACCTCCCCTGGCACCACAGTCTTGTGTTTTTGGCACCAGCAAAGATATTGGCACAAA
TGCTGAGATCACTTATCTCATCCGGTCTGGGAACGAACAAGGGAAATTTAAGATCAACCC
CAAGACAGGGGGTATTTCTGTCTCTGAAGTCCCTGGACTATGAATTATGCAAAAGGTTTTA
CCTGGTAGTGGAAGCCAAAGATGGGGGCACCCAGCTCTCAGCGCTGTGGCCACTGTCAA
CATCAACCTCACAGATGTTAATGACAACCTCCCAAGTTCAGCCAAGACGTCTACAGTGC
GGTTATCAGTGAAGACGCCTTGGTGGGAGACTCTGTCAATTTTGCTAATAGCAGAAGATGT
AGACAGCCAGCCCAACGGACAGATTCAATTTTCCATTGTGAATGGAGATCGGGACAATGA
ATTACTGTAGATCCTGTCTTGGGACTTGTGAAAGTTAAGAAGAAATTGGACCGGGAACG
GGTGTCTGGATACTCTCTGCTTGTCCAGGCCGTAGACAGTGGCATTCTGCAATGTATC
AACTGCAACTGTCAACATTGATATTTCTGATGTGAATGACAACAGCCCGGTGTTTACACC
TGCCAACCTATACTGCTGTGATTTCAGGAAAATAAGCCAGTGGGCACCAGCATCTTGCAGCT
GGTGGTGACAGACAGAGACTCCTTTTACAATGGGCCTCCCTTTTCACTCTATTTTGTG
GGGAAATGAAGAGGAGGAGTTTGTGTTGGACCCCTCATGGGATCTTGGCGTCCGGCTGTGGT
CTTCACGACACAGAGTCTCTGGAATACGTGTTGTGTGTCCAGGCAAAGGATTCAGGCCAA
ACCCAGCAAGTTTCTCACACTTACATCCGCGTGCAGATCATTGAGGAAAGCACCCACAA
GCCACAGCCATTCCCTGGAAATTTTCACTTGTCAACATGGAGGATGACTTTCCTGGTGG
GGTCATTGGGAAGATTTCATGCCACAGATCAAGACATGTATGATGTGCTCACATTTGCCCT
GAAATCGGAGCAGAAAAGCTTATTTAAAGTGAACAGTCACGATGGGAAAATCATCGCCCT
GGGAGGCCTGGACAGCGCAAGTATGTCTGAATGTGTCTGTGAGTGATGGTTCGCTTCCA
GGTACCCATTGATGTGGTTCGTGCATGTGGAGCAGTTGGTGCATGAGATGTGTCAGAACAC
TGTCACCATCCGCTTTGAAAATGTGTCCCTGAGGACTTCGTGGGGCTGCACATGCATGG
GTCCCGGCGCACCCCTGCGGAATGCAGTCTCACCCAGAACGAGACAGCCTGCGCATCAT
CAGCATCCAGCCCGTGGCAGGCACCAACCAACTGGACATGCTGTTTGGCGTGGAGATGCA
CAGCAGCGAGTTCTACAAGCCAGCCTACCTGATCCAGAAGCTGTCCAATGCTAGAAGACA
CCTGGAGAATATCATGCGCATCTCAGCCATCTTGGAGAAGAACTGCTCAGGGCTGGACTG
TCAGGAACAGCATTGTGAGCAAGGCTTGTCACTCGATTCCACGCGCTCATGACCTACAG
CACGGCTCGCATCAGCTTTGTGTGTCCGCGTTTCTACAGGAACGTGCGTTGCACCTGCAA
TGGTGGACTGTGTCCGGGGTCCAACGATCCTTGTGTGGAGAAGCCGTGTCCAGGGGACAT
GCAGTGTGTGCTAGTTATGAAGCCAGCAGGAGACCGTTCCCTCTGCCAGTGTCCACGAGGAA
GCTCGGAGAGTGTCTCAGGGCACACTTCTCTCAGCTTTGCTGGAACAGTTACATCAAATA
TCGGCTTTTCTGAAAATAGCAAAGAAGAGGATTTCAAACCTAGCTCTGCGTCTTCAACACT
GCAAAGCAATGGGATTATAATGTACACCAGAGCAAATCCCTGCATAATTCTGAAGCAGAT
TGTGGATGGCAAGCTGTGGTTCCAGCTGGACTGCGGCAGCGGCCCTGGAATCTTGGGCAT
CTCGGGCCGTGCTGTCAACGACGGGAGCTGGCACTCGGTCTTCTGGAGCTCAACCGCAA
TTTCACGAGCCTGTCCCTGGATGACAGCTACGTGGAGCGGCGCCGGGCGCCCTCTACTT
CCAGACGCTGAGCACTGAGAGTAGCATCTACTTCGGCGCCCTGGTGCAAGCGGATAACAT
CCGACGCTGACTGACACGCGGGTCACGCAGGTGCTCAGCGGCTTCCAGGGCTGCCTGGA
CTCGGTGATACTGAATAACAATGAGCTGCCGCTGCAGAACAAGCGCAGCAGCTTCGCGGA
GGTGGTGGGCTGACGAGCTGAAGCTGGGCTGCGTGTCTATCCCGACGCTGCAAGCG
CAGCCCGTGCCAGCACGGGGGAGCTGCACTGGCCTGCCATCGGGGGTGGCTATCAGTG
TACCTGTCTCTCACAGTTTACGGGGAGAACTGTGAATCTGAGATTACAGCCTGCTTCCC
AAACCCCTGCCGGAATGGAGGATCCTGCGATCCAATAGGAAACACTTTCATCTGCAATTG
TAAAGCTGGGCTCACTGGAGTCAGTGTGAGGAGGACATCAATGAGTGCGAACGAGAGGA
GTGTGAGAACGGAGGCTCCTGCGTGAACGTGTTCCGGCTCCTTCTCTGCAACTGCACGCC
GGGCTACGTGGGCCAGTACTGCGGGCTGCGCCCCGTGGTGGTACCCAATATCCAGGCTGG
CCACTCCTACGTGGGGAAGGAGGAGCTCATCGGCATCGCCGTGGTCTCTTCTGTATCTT
CATCCTGGTGGTTCTTTCATAGTCTTCCGCAAGAAGGTCTTCCGCAAGAATACTCCCG
CAACAACATCACGCTAGTGACAGGACCCGGCCACCGCCGCTGCTTAACAAGAGCAATGG
CATCCCGTTCCGGAACCTGCGCGGCAGTGGGGACGGCCGCAACGTCTACCAGGAGGTGGG
GCCCCCGCAGTCCCCGTGCGCCCCATGGCTACACACCTGCTTCCAGAGTGACTCCAG

GAGCAACCTGGATAAGATCGTGGACGGGCTGGGAGGCGAGCACCAGGAAATGACCACGTT
 TCACCCTGAGTCGCCCCGCATCCTGACAGCCCGGCGGGCGTGGTTCGTGTGCAGTGTGGC
 CCCCACCTCCCCGCGTGTACCCCTGCCGCTCCGACTGCGACTCCATCCGGAAGAATGG
 CTGGGACGCGGGAACGTAGAGTAATAAAGGCAGCAACTCTGAAGTTCAGTCCCTCAGCTC
 CTTCAGTCAGATTCTGGTGACGACAATGCCTATCACTGGGACACCTCTGATTGGATGCC
 AGGGGCGCGCTGTCTGGACATAGAGGAAGTGCCCAACTATGAGAACCAGGATGGAGGGTC
 TGCACACCAGGGGAGCACACGGGAGCTGGAGAGCGATTACTACCTGGGTGGTTATGACAT
 TGACAGTGAATACCCACCCCTCATGAAGAGGAGTTCTTGAGTCAGGACCAGCTGCCTCC
 TCCTCTCCCGGAGGACTTCCAGACCAATATGAGGCCCTGCCACCCTCCCAGCCTGTCTC
 CCTGGCCAGCACACTGAGCCAGACTGCAGGAGAAGGCCCCAGTTTCATCCTAGCCAGTA
 TCTCCCTCCTCACCCATTCCCCAACGAAACGGATTGGTGGGCGCGCTGCCAGCTGTGA
 ATTTAGTACTTTTGTGTGAGCATGAACCAGGGCACAGAGCCACAGGCCACAGGCCAGCAGAG
 CGTGTCTCTGTCTTGCACAATTCCAGAGGCACCTCATCCTCGGATGTGTCTGCCAACTG
 CGGCTTTGACGATTCCGAAGTAGCCATGAGTGACTACGAGAGCGTGGGAGAGCTCAGCCT
 CGCCAGCCTTACATTCCCTTTGTGGAGACTCAGCATCAGACTCAAGTGTAGACATCACA
 TCTTGGGTACTTCACCCTGT

In a search of public sequence databases, the NOV9 nucleic acid sequence, located on
 chromosome 11 has 4976 of 7882 bases (63%) identical to a gb:GENBANK-
 ID:AF100960|acc:AF100960.1 mRNA from Rattus norvegicus (Rattus norvegicus
 5 protocadherin (Fat) mRNA, complete cds). Public nucleotide databases include all GenBank
 databases and the GeneSeq patent database.

The disclosed NOV9 polypeptide (SEQ ID NO:22) encoded by SEQ ID NO:21 has
 4544 amino acid residues and is presented in Table 9B using the one-letter amino acid code.
 Signal P, Psort and/or Hydropathy results predict that NOV9 has 2 signal peptide and is likely
 10 to be localized extracellularly with a certainty of 0.4600. The most likely cleavage site is
 between residues 32 and 33.

Table 9B. Encoded NOV9 protein sequence (SEQ ID NO:22).

MDIIMGHCVGTRPPACCLILLFLKLLATVSQGLPGTGLGFHFTHSIYNATVYENSAART
 YVNSQSRMGITLIDLSDIKYRIVSGDEEGFFKAEEV IADFCFLRIRTKGNSAILNRE
 IQDNYLLIVKGSVRGEDLEAWTKVNIQVLDMDLRLPLFSPTTYSVTIAESTPLRTSVAQV
 TATDADIGSNGEFYYYFKNKVDLFSVHPTSGVISLSGRLNYDEKNRYDLEILAVDRGMKL
 YGNNVGSSTAKLYVHIERINEHPTIHVVTHVPFSLEKEPTYAVVTVDLDDGANGEIES
 VSIVAGDPLDQFFLAKEGKWLNEYKIKERKQIDWESFPYGYNLTLQAKDKGSPQKCSALK
 AVYIGNPTRDTPIRFEKEVYDVSISEFSPPGVVVAIVKLSPEPIDVEYKLSPEGDAVYF
 KINPRSLIVTARPLNTVKKEVYKLEVTNKEGDLKAQVTISIEDANDHTPEFQQPLYDAY
 VNESVPVGTSVLTVSASDKDKGENGYITYSIASLNLPLPFVINQFTGVISTTEELDFESSP
 EIYRFIVRASDWGSPYRHESEVNVITIRIGNVNDNSPLFEKVACQGVISYDFPVGGHITAV
 SAIDIDELELVKYKII SGNELGFFYLNPD SGVLQLKKS LTN SGIKNGNFALRITATDGEN
 LADPMSINISVLHGKVSSKSFSCRETRVAQKLAEKLLIKAKANGKLNLEDGFLDFYSINR
 QGPYFDKSFPSDVAVKEDLPVGANILKIKAYDADSGFNGKVLFTISDGN TDS CFNIDMET
 GQLKVLMPMDREHTDLYLLNITIYDLGNPQKSSWRLLTINVEDANDNSPVFIQDSYSVNI
 LESSGIGTEIIQVEARDKDLGSNGEVTYSVLTDTQQFAINSSTGIVYVADQLDRESKANY
 SLKIEARDKAESGQQLFSVVT LKVFLDDVND CSPA FIPSSYSVKVLEDLPVGTVI AWLET
 HD PDLGLGGQVRYSLVNDYNGRFEIDKASGAIRLSKELDYEQQFYNLTVRAKDGRPV S
 LSSVSFVEVEVVDVNENLHTPYFPDFAVVGSVKENSRIGHTSVLQVTARDEDSGRDGEIQY
 SIRDGSLGRFSIDDESGVITAADILDRETMSYWLTVYATDRGVVPLYSTIEVYIEVED
 VNDNAPLTSEPLYIILVMDKHPKDVSVI IQIQAEDPDSSSNEKLT YRITSGNPQNFLCINI
 KTGLITTTSRKLDREQQA EHFLEVTVTDGGPSPKQSTI WVVVQVLDENDNKPQFPEKVYQ

```

IKLPERDRKKRGEPIYRAFAFDRDEGPNAEISYSIVDGNDDGKFFIDPKTGMVSSRKQFT
AGSYDILTIAVDNRPQKSSTARLHIEWIKKPPSPPIPLTFDEPFYNTVMESDRVTEI
VGVVSVQPANTPLWFDIVGGNFDSAFDAEKVGTVIAKPLDAEQRSIYNMSVEVTDGTN
VAVTQVFIKVLNNDNGPEFSQPNYDVTISEDVLPDTEILQIEATDRDEKHKLSYTVHSS
IDSISMRKFRIDPSTGVLYTAERLDHEAQDKHILNIMVRDQEFYRRNLARVIVNVEDAN
DHSPTYFTNPLYEASVFESAALGSAVLQVTALDKDKGENAELIYTI EAGNTGNMFKIEPVL
GIITICKEPDMTTMGQFVLSIKVTDQGSPPMSATAIVRISVTMSDNSHPKFIHKDYQAEV
NENVDIGTSVILISAIQSSTLIYEVKGDINGIFTINPYSGVITTQKALDYERTSSYQLI
IQATNMAGMASNATVNIQIVDENDNAPVFLFSQYSGSLSEAAPINSIVRSLDNSPLVIRA
TDADSNRNALLVYQIVESTAKKFFTVDSSSTGAIRTIANDHETIAHFHFHVHRDMSGSPQ
LTAESPVEVNIETDVNDNPPVFTQAVFETILLPTVVGVEVLKVSATDPDSEVPPELTY
SLMEGSLDHFLIDSNSGVLTIKNNNSKDHVMLIVKVS DGKFYSTSMVTIMVKEAMDSGL
HFTQSFYSTSISENNTNITKVAIVNAVGNRLNEPLKYSILNPGNKFKIKSTSGVIQTTGV
PFDREEQELYELVVEASRELDHLRVARVVVRVNI EDINDNSPVFVGLPYAAVQVDAEPG
TLIYQVTAIDKDKGPNGEVTYVLQDDYGHFEINPN SGNVILKEAFNSDLSNIYGVITILA
KDGGKPSLSTSVELPITIVNKAMPVFDKPFYTA SVNEDIRMNTPILSINATSPGQGIIY
IIIDGDPFKQFNIDFDTGVLKVVSPLDYEVTSAYKLTIRASDALTGARAEVTVDLLVNDV
NDNPPIFDQPTYNTTLSEASLIGTPVLQVVSIDADSENNKMVHYQIVQDTYNSTDYFHID
SSSGLILTARMLDHELVOHCTLKVRSIDSGFP SLSSEVLVHIYISDVNDNPPVFNQLIYE
SYVSELAPRGHFVTCVQASDADSSDFDRLEYS ILSGNDRTSFLMDSKSGVITLSNHRKQR
MEPLYSLVSVSDGLFTSTAQVHIRVLGANLYSPA FSQSTYVAEVRENVAAGTKVIHVRA
TDGDPGTYGQISYAIINDFAKDRFLIDSNGQVIT TERLDRENPLEGDVSI FVRALDGGGR
TTFCTVRVIVVDENDNAPQFMTVEYRASVRADV GRGHLVTQVQAIDPDDGANSRITYSLY
SEASVSVADLLEIDPDNGWMVTKGNFNQLKNTVLS FFFVKAVDGGIPVKHSLIPVYIHVLP
PETFLPSFTQSQYSFTIAEDTAIGSTVDTLRILPS QNVWFSTVNGERPENNKGGFVIEQ
ETGTIKLDRKLDRETSAPHFVKVAATIPLDKVDIVFTVDVDIKVLDLNDNKPVFETSSYD
TIIMEGMPVGTCLTQVRAIDMDWGANGQVTYSLHSD SQPEKVM EAFNIDSNTGWISTLKD
LDHETDPTFTFSVASDLGEAFSLSSSTALVSVR VTDINDNAPVFAQEVYRGNVKESDPPG
EVVAVLSTWDRDTSVNRQVSYHITGGNPRGRFALGLVQ SEWKVYVKRPLDREEQDIYFL
NITATDGLFVTQAMVEVSVSDVNDNSPVC DQVAYTALLPEDIPSNKIILKVS AKDADIGS
NGYIRYSLYGSGNSEFFLDPESGELKTLALLDRERIPVYSLMAKATDGGGRFCQSNHILI
LEDVNDNPPVFSSDHYNTCVYENTATKALLTRVQAVDPDIGINRKVVYSLADSAGGVFSI
DSSSGIIILEQPLDREQQSSYNISVRATDQSPGQSL SSSLTTVTITVLDINDNPPVFERRD
YLVTVPEMTSGPTQVLAVFATSKDIGTNAEITYLIRSGNEQ GKFKINPKTGGISVSEVLD
YELCKRFYLVVEAKDGGTPALSAVATVNIINLTDVNDNPPKFSQDVYSAVISEDALVGDSV
ILLIAEDVDSQPNGQIHFSIVNGDRDNEFTVDPVLGLVKVKKKLDRE RVS GYSLLVQAVD
SGIPAMSSTATVNIDISDVNDNSPVFT PANYTAVIQENKPVGTSILQLVVTDRDSFHNGP
PFSFISILSGNEEEFVLDPHGILRS AVVFQHTESLEYVLCVQAKDSGKPKQVSHTYIRVR
VIEESTHKPTAIPLEIFIVTMEDDFPGGVIGKIHATDQDMYDVLTFALKSEQKSLFKVNS
FDGKIIALGGLDSGKYVLNVSVSDGRFQVPIDVVHV EQLVHEMLQNTVTIRFENVSPED
PVGLHMHGFRRTL RNAVLTFQKQDSLRIISIQPVAGTNQLDMLFAVEMHSSEFPKPAYLIQ
KLSNARRHLENIMRISAILEKNC SGLDCQE QHCEQGLSLDSHALMTYSTARISFVCPRFY
RNVRCTCNGGLCPGSNDPCVEKPCPGDMQCVSYEASRRPFLCQCPPGKLGECSGHTSLSF
AGNSYIKYRLSENSKEEDFKLALRLRTLQSNGIIMYTRANPCIILKQIVDGKLFQWLD CG
SGPGILGISGRAVNDGSWHSVFLELNRNFTSLSLD DS YVERRRAPLYFQTLSTESSIYFG
ALVQADNIRSLTDTRVTQVLSGFQGC LDSVILNNNELPLQNKRSSFAEVVGLTELKLGCV
LYPDACKRSPCQHGGSC TGLPSGGGYQCTCLSQFTGRNCESEITACFPNPCRNGGSCDPI
GNTFICNCKAGLTGVTCEEDINECERECEENGGS CVNVFGSFLCNCTPGYVGQYCGLRPV
VVPNIQAGHSYVGKEELIGIAVVL FVIFILVVL FIVFRKKVFRKNYSRNNITLVQDPATA
ALLNKSNGIPFRNLRGSGDGRNVYQEVGPPQVPVRPMAYTPCFQSDSRSNLDKIVDGLGG
EHQEMTTTFHPESPRILTARRGVVVC SVAPNLP AVSPCRSDCDSIRKNGWDAGTESNKGSN
SEVQSLSSSFQSDSGDDNAYHWDTS DWMPGARLSDIEEVPNYENQDGGS AHQGSTRELES D
YYLGGYDIDSEYPPPHHEEFLSQDQLPPPLPEDFPDQYEALPPS QPVSLASTLSPDCRRR
PQFHPSQYLPHPFPNETDLVGPPASCE FSTFAVSMNQGT EPTGPADSVSLSLHNSRGTS
SSDVSANCGFDDSEVAMS DYESVGELSLASLHIPFVETQHQTQV

```

A search of sequence databases reveals that the NOV9 amino acid sequence has 2201 of 4118 amino acid residues (53%) identical to, and 2931 of 4118 amino acid residues (71%) similar to, the 4589 amino acid residue ptnr: SPTREMBL-ACC:Q9WU10 protein from *Rattus*

norvegicus (Rat) (PROTOCADHERIN). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV9 is expressed in at least breast, prostate, bone marrow, brain, liver, stomach, pituitary, cartilage. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV9 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 9C.

Table 9C. BLAST results for NOV9					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 18578022 ref XP_061871.2 (XM_061871)	similar to FAT tumor suppressor (Drosophila)	3446	3319/3386 (98%)	3320/3386 (98%)	0.0
gi 4885229 ref NP_05236.1 (NM_005245)	FAT tumor suppressor precursor	4620	2381/4620 (51%)	3178/4620 (68%)	0.0
gi 13929168 ref NP_114007.1 (NM_031819)	FAT tumor suppressor (Drosophila) homolog [Rattus norvegicus]	4589	2368/4619 (51%)	3170/4619 (68%)	0.0
gi 6688786 emb CAB65271.1 (AJ250768)	mouse fat 1 cadherin [Mus musculus]	4587	2367/4633 (51%)	3167/4633 (68%)	0.0
gi 13787217 ref NP_001438.1 (NM_001447)	FAT tumor suppressor 2 precursor [Homo sapiens]	4349	1854/4434 (41%)	2635/4434 (58%)	0.0

Table 9D lists the domain descriptions from DOMAIN analysis results against NOV9. This indicates that the NOV9 sequence has properties similar to those of other proteins known to contain this domain.

Table 9D. Domain Analysis of NOV9
<p>gnl Smart smart00112, CA, Cadherin repeats.; Cadherins are glycoproteins involved in Ca²⁺-mediated cell-cell adhesion. Cadherin domains occur as repeats in the extracellular regions which are thought to mediate cell-cell contact when bound to calcium.</p> <p>CD-Length = 82 residues, 100.0% aligned</p> <p>Score = 97.4 bits (241), Expect = 2e-20</p>

Table 9E. Domain Analysis of NOV9

gnl|Pfam|pfam00028, cadherin, Cadherin domain
 CD-Length = 92 residues, 100.0% aligned
 Score = 94.4 bits (233), Expect = 1e-19

NOV9 is a member of the protocadherin family, which in turn is one of the six subfamilies of the cadherin superfamily. Cadherins are membrane-associated glycoproteins that mediate cell-cell interactions in a calcium-dependent fashion. Protocadherins may act as cell-cell recognition molecules and may be involved in signal transduction cascades.

NOV9 has homology to the rat protocadherin that is most related to the *Drosophila* FAT gene. The *Drosophila* FAT gene shows the presence of multiple characteristic cadherin domains and is likely involved in cell guidance, cell repulsion and/or cell adhesion. Recessive lethal mutations in the fat locus of *Drosophila* cause hyperplastic, tumor-like overgrowth of larval imaginal discs, defects in differentiation and morphogenesis, and death during the pupal stage. This indicates that the fat gene has a tumor suppressor function (See Mahoney et al., *Cell* 1991 Nov 29;67(5):853-68).

The disclosed NOV nucleic acid of the invention encoding a protocadherin-like protein includes the nucleic acid whose sequence is provided in Table 9A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 9A while still encoding a protein that maintains its protocadherin-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 37 percent of the bases may be so changed.

The disclosed NOV9 protein of the invention includes the protocadherin-like protein whose sequence is provided in Table 9B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table

9B while still encoding a protein that maintains its protocadherin-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 47 percent of the residues may be so changed.

5 The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this protocadherin-like protein (NOV9) may function as a member of a "protocadherin family". Therefore, the NOV9 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

15 The NOV9 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the protocadherin-like protein (NOV9) may be useful in gene therapy, and the protocadherin-like protein (NOV9) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, graft versus host disease, endocrine dysfunctions, diabetes, obesity, growth and reproductive disorders, Von Hippel-Lindau (VHL) syndrome, cirrhosis, transplantation, hypercalcaemia, ulcers, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neurodegeneration, cancer, tissue degeneration, bacterial/viral/parasitic infections, or other pathologies or conditions. The NOV9 nucleic acid encoding the

20

25

30

protocadherin-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV9 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV9 substances for use in therapeutic or

diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV9 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV10

A disclosed NOV10 nucleic acid of 1071 nucleotides (also referred to as CG57553-01) encoding a T01C1.3-like protein is shown in Table 10A. Putative untranslated regions upstream and/or downstream from the coding region, if any, are underlined, and the start and stop codons are in bold letters.

Table 10A. NOV10 nucleotide sequence (SEQ ID NO:23).	
<p>ATGCATCAGTTTACCATCGTTTCATTCCATCCATCCAGGAGACTACACAGAAACAGAGAG GACTATGTGGAAAGAAGTGCTGAGTTTGAGATGGTTTGCTCTCAAAGCTTTGAAAGAC ATTCAGTCTGGAGCACTGGACATAAATAAAGCAGGCATACTTTATGGCATACTCAGAGAG ACTTTACTTCTTCACTTAGAAGCCTTACCAGCAGGGAAGCCTGCATCTTTTAAAAACAAA ACTCGAGATTTCCATGATAGTTATTTCATATAAGGACAGTAAAGAACTTGTGCAGTGCTG CAAAAAGTAGCCTTGTGGGCAAGAGCTCAAGCAGAGCGCACAGAAAAAAGTAACTCAAT CTACTTGAAACCTCAGAAATAAAATTCCCAACAGCTTCCACTTACCTCCATCAGCTAACT CTACAGAAAATGGTCACTCAGTTTAAAGAAAAAATGAAAGCCTCCAATATGAAACTTCA AATCCTACTGTACAGTTAAAAATTCTCAGCTACGAGTAAGTTCTGTCTCAAAATCACAA CCTGATGGTTCTGGTCTGTTGGATGTTATGTATCAAGTTTCCAAAACCTCTTCAGTCCTA GAAGGATCAGCTCTCCAAAACCTGAAAAATATACTCCCTAAACAGAACAAAATAGAATGT TCTGGGCCTGTAATCACTCAAGTGTGACTCTTACTTTCTACATGGGGACCTCTCTCCT TTGTGTCTTAATTTCTAAAAATGGAACAGTTGATGGAACCTCTGAAAATACTGAAGATGGA TTAGATCGAAAAGACAGTAAGCAGCCAGGAAAAAACGTGGCCGCTATCGGCAATATGAT CATGAAATAATGGAAGAAGCTATTGCAATGGTAATGAGCGGAAAAATGAGTGTTCCTCAA GCACAAGGAATTTATGGGGTACCTCACAGCACTTTAGAATACAAGGTAAGAAAGATCT GGAACACTGAAGACTCCTCCGAAGAAGAACTACGATTACCAGACACTGGGTTATATAAT ATGACAGATTTCAGGACTGGCAGCTGCAAAAACAGCAGCAAGCCTGTGTAG</p>	

In a search of public sequence databases, the NOV10 nucleic acid sequence, located on chromosome 4 has 165 of 267 bases (61%) identical to a gb:GENBANK- ID:BGDNA66KD|acc:X87727.1 mRNA from *Borrelia garinii* (B.garinii p66 gene for 66kDa protein). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV10 polypeptide (SEQ ID NO:24) encoded by SEQ ID NO:23 has 356 amino acid residues and is presented in Table 10B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV10 has no signal peptide and is likely to be localized in the nucleus with a certainty of 0.7000.

Table 10B. Encoded NOV10 protein sequence (SEQ ID NO:24).
MHQFTIVSFHPSRRLHRNREDYVERSAEFADGLLSKALKDIQSGALDINKAGILYGIPIQK TLLHLEALPAGKPASFKNKTRDFHDSYSYKDSKETCAVLQKVALWARAQAERTEKSKLN LLETSEIKFPTASTYLHLQTLQKMVTQFKEKNESLQYETSNPTVQLKIPQLRVSSVSKSQ PDGSGLLDVMYQVSKTSSVLEGSALQKLKNILPKQNKIECSGPVTHSSVDSYFLHGDLS LCLNSKNGTVDGTSENTEDGLDRKDSKQPRKKRGRYRQYDHEIMEEAIAMVMMSGKMSVSK AQGIYGVPHSTLEYKVKERSGTLKTPPKKKLRLPDTGLYNMTDSGTGSCKNSSKPV

5

10

15

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 18559097 ref XP_087581.1 (XM_087581)	protein XP_087581 [Homo sapiens]	213	213/213 (100%)	213/213 (100%)	e-110
gi 16549953 dbj BAB70892.1 (AK055258)	unnamed protein product [Homo sapiens]	213	212/213 (99%)	213/213 (100%)	e-109
gi 14017807 dbj BAB47424.1 (AB058698)	KIAA1795 protein [Homo sapiens]	572	94/242 (38%)	125/242 (50%)	7e-28
gi 14744761 ref XP_050988.1 (XM_050988)	KIAA1795 protein [Homo sapiens]	433	94/242 (38%)	125/242 (50%)	4e-26
gi 18487927 ref XP_081696.1 (XM_081696)	Eip93F [Drosophila melanogaster]	1165	37/76 (48%)	46/76 (59%)	6e-10

15

metabolic disorders. The T01C1.3-like gene maps to human chromosome 4 and has no identifiable domains.

The disclosed NOV10 nucleic acid of the invention encoding a T01C1.3-like protein includes the nucleic acid whose sequence is provided in Table 10A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 10A while still encoding a protein that maintains its T01C1.3-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 39 percent of the bases may be so changed.

The disclosed NOV10 protein of the invention includes the T01C1.3-like protein whose sequence is provided in Table 10B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 10B while still encoding a protein that maintains its T01C1.3-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 68 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$ that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this T01C1.3-like protein (NOV10) may function as a member of a “T01C1.3 family”. Therefore, the NOV10 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV10 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the T01C1.3-like protein (NOV10) may be useful in gene therapy, and the T01C1.3-like protein (NOV10) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from cancer, trauma, bacterial and viral infections, in vitro and in vivo regeneration, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neurodegeneration, diabetes, autoimmune disease, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, and hypercalcaemia, or other pathologies or conditions. The NOV10 nucleic acid encoding the T01C1.3-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV10 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV10 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV10 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV 11

NOV11 includes three alpha-macroglobulin-like proteins disclosed below. The disclosed sequences have been named NOV11a, NOV11b and NOV11c.

NOV 11a

A disclosed NOV11a nucleic acid of 6195 nucleotides (also referred to as CG57488_01) encoding a alpha-macroglobulin-like protein is shown in Table 11A. Putative untranslated regions upstream and/or downstream from the coding region, if any, are underlined, and the start and stop codons are in bold letters.

Table 11A. NOV11a nucleotide sequence (SEQ ID NO:25).

ATGAGCGGCGCCCTGCTCTGGCCGTTGCTCCCGCTCCTGCTCCTGCTGCTGCTGCGGCGCGG
GACGGCGTGCGCGCCGCGCAGCCTCAGGCCCCGGGTACTTGATTGCAGCTCCCTCTGTT
TTTCGCGCGGGCGTGAGGAAGTCATCAGCGTGACCATCTTTAACTCTCCAAGGGAAGTC
ACGGTCCAGGCTCAGCTGGTGGCCAGGGTGAGCCGGTGGTGCAGAGCCAGGGAGCCATC
CTGGATAAAGGGACAATCAAATCAAGCATACGGTCTCAGCACCTCCGGTATCTCCCTC
CTGCCCATCCTTGCCCTGCTCTTGGGTGGCCGGGACCTTTCCTCCCTCTTCAGCTCTGG
CCAGTGTGAGATATTTCCAGAAACAGGGCCAGGTGCCCACGGGCCTCCGGGGCCAAGCG
CTTCTGAAAGTGTGGGGCCGCGGCTGGCAGGCGGAGGAGGGGCCCTCTTTCACAACCAG
ACCTCGGTGACCGTGGACGGCCGGGGCGCTTCTGTATTATCCAGACGGACAAGCCTGTG
TACAGACCCAGCAGCAGGTGCTCATAAGCATCTTCACCGTCTCTCCAAATCTGAGGCCT
GTCAACGAGAAGCTGGAAGCCTACATCCTGGACCCCCGAGGCTCTCGGATGATAGAGTGG
AGACACTTAAAGCCGTTCTGCTGCGGCATCACCAACATGAGCTTCCCCTTGTCGACCAG
CCTGTGTTGGGAGAATGGTTCATTTTTGTTGAAATGCAAGGCCACGCGTACAACAAGTCT
TTTGAAGTTTCAAGATATGTTTGCCCAAGTTTGAGCTTCTGATTGACCCGCCCCGGTAT
ATCCAAGACCTGGACGCTGTGAGACAGGCACTGTGCGGGCCAGGTATACCTTTGGGAAA
CCTGTGGCTGGTGCCTTAATGATCAACATGACTGTTAATGGTGTAGGGTACTACAGCCAC
GAGGTGGGACGCGCTGTCTCAGAACCAAGATCCTCGGCTCCCGGGACTTCGACATC
TGCGTGAGGGACATGATCCAGCGGACGTCCCTGAGCACTTCCGGGGCAGGGTCAGCATC
TGGGCCATGGTGACAGTGTGGACGGGAGCCAGCAGGTGCGGTTGATGACTCCACCCCC
GTGCAGAGGCAGCTGGTGGACATCCGGTACTCCAAGGACACGAGGAAGCAGTTCAAGCCG
GGCCTGGCCTACGTGGGGAAGGTGGAGCTATCCTACCCCGATGGCAGCCACGTGAGGGG
GTGACGGTCCAGATTAAGGCAGAGCTGACACCAAAGGATAACATCTACACCAAGTGAAGTT
GTGTCCCAGCGTGGACTAGTGGGGTTTGAATCCCTCCATCCCCACGTGAGCCAGCAC
GTGTGGCTGGAGACCAAGGTGATGGCACTGAACGGGAAGCCCGTGGGGGCTCAGTACCTG
CCAGCTACCTCTCCCTCGGCAGCTGGTACTCCCCCAGCCAGTGCTACCTGCAGCTGCAG
CCACCCTCCACCCACTGCAGGTTGGGGAAGAAGCCTATTTTTCTGTGAAGTCCACATGT
CCCTGCAACTTTACCCTGTACTACGAGGTGGCTGCACGGGGCAATATTGTGTATCGGGC
CAGCAGCCTGCCCACACCACCCAGCAGCGAAGCAAGCGGGCGGCCCCCTGCCCTGGAGAAA
CCGATTCTGTTTAAACACACCTTTCTGAGACAGAGCCCCCACCAGCCCCAGAAGCTGAGGTC
GACGTGTGTGTGACCTCTCTCATCTGGCCGTGACCCCCAGCATGGTCCCCCTTGGTCGC
GTGTGGCTCTTCTACGTGAGGGAATGGAGAAGGGGTGCGCCGACAGCCTTCAAGTTGCA
GTGAGACCTTCTTCGAAAACAGGTTTTCAGTGACGTATTCAGCAAATGAGACCCAACCT
GGGGAGGTTGTGACCTGCGGATCAGGGCTGCAAGGGGCAGCTGTGTGTGCGTCGCGCA
GTTGATAAGAGTGTCTACCTGCTCAGGTCTGGGTTCGGCTGACTCCTGCCAGGTTTTC
CAGGAAGTGAAGATTATGATGTTTCTGATTCTTTGGCGTGTCCAGGGAGGATGGTCCT
TTTTGGTGGGCTGGGCTGACGGCACAACGACGCCGGCGCTCCTCTGTCTTCCCGTGGCCT
TGGGGCATCACCAAGGACTCTGGGTTTGCTTACCGAAACGGGACTGGTGGTGATGACC
GACCGAGTGAGCTGAACACCGGCAGGACGGTGGCTCTACACCGATGAGGCTGTCCCC
GCTTTTCAGCCCCACACAGGGACCTGGTGGCAGTGGCTCCTTTCAGGCACCCCCCAGA
ACAGAGAAGAGAAAAGGACTTTCTTCCCCGAAACATGGATTGTCATTGTCTCAACATC
AGTGACCCATCTGGTGAGGGGACACTCAGTGTGAAGGTCCCGGACTCCATCACCAGCTGG
GTGGGTGAGGCCGTGGCCCTGTCCACCTCTCAGGGCTTAGGCATCGCCGAGCCCTCCCTG
CTGAAGACCTTCAAGCCCTTCTTCGTGGACTTCATGCTCCCCGCTCTCATCATCCGTGGG
GAGCAGGTCAAGATCCCGCTCAGTGTCTACAACATACATGGGCACCTGCGCTGAGGTGTAC
ATGAAGCTCTCGGTTCCCAAGGGCATCCAGTTTGTGGGCATCCTGGCAAACGCCATGTG
ACCAAGAAGATGTGTGTGGCCCCCGGGAGGCTGAGCCCATCTGGGTCTTCTGTCTCTTC
AGCGACCTGGGACTCAACAACATCACGGCCAAAGCCCTTGCTTACGGAGACACAAATTGC
TGCCGGGATGGGAGGTCCAGCAAACACCCTGAGGAGAATCACGCCGACAGGAGGGTCCCC
ATCGGGGTGGATCACGTACGGCGCAGTGTGATGGTTGAGGCGGAAGGATCCCCGGGCG
TACACCTACAGCGATTCTTCTGTCCAGTGAGAGAGTCCACATCTCCACCCCCAACAAAG
TATGAGTTCCAGTATGTGCAGCGGCCACTGCGCCTCACCCGCTTTGATGTGGCTGTGCGA
GCTCACAATGATGCCCCGTGTGGCCTTGTCTTCTGGGCCCCAGGACACAGCAGGATGATC
GAGATCGTCTGGGGGGCATCAGAACACAGGTATGGATCTCCACGCAAGATGGA
GAGCCCGTGGCCAGTGACACACGGCCAAGATCCTCTCTGGGATGAATTCAGAACATTC
TGGATCAGCTGGCGTGGTGGCCTTATCCAGGTTGGCCATGGTCCAGAGCCATCCAATGAG
TCTGTCAATTGTGGCCTGGACCCTCCCAGGGCCACCAGAGGTCCAGTTCATTGGCTTTTCC
ACCGGCTGGGGCTCCATGGGTGAATTCCGAATCTGGAGGAAGATGGAGGTGGACGAGAGC
TACAGCGAGGCCTTACCCTGGGGGTCCCACACGGCGCCATCCCTGGGTCTGAGCGAGCC

[illegible]

5

The disclosed NOV11a polypeptide (SEQ ID NO:26) encoded by SEQ ID NO:25 has 1927 amino acid residues and is presented in Table 11B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV11a has a signal peptide and is

likely to be localized at the plasma membrane with a certainty of 0.6400. The most likely cleavage site for a NOV11a peptide is between amino acids 25 and 26.

Table 11B. Encoded NOV11a protein sequence (SEQ ID NO:26).

MSGALLWPLLPLLLLLLLSARDGVRAAQQPAPGYLIAAPSVFRAGVEEVISVTIFNSPREV
TVQAQLVAQGEPPVQSQGAILDKGTIKLKHTVLSTSGISLLPILALLLGGRLDSSLFSLW
PVLRYFQKQGQVPTGLRGQALLKVWGRGWQAEEGPLFHNQTSVTVDGRGASVFIQTDKPV
YRPQHRVLISIFTVSPNLRPVNEKLEAYILDPRGSRMIEWRHLKPFCCGITNMSFPLSDQ
PVLGEWFI FVEMQGHAYNKSFEVQKYVLPKFELLIDPPRYIQDLDACETGTVRARIYTFGK
PVAGALMINMTVNGVGYGSHEVGRVPLRTTKILGSRDFDICVRDMIADVPHEHFRGRVS
WAMVTSVDGSSQVAFDDSTPVQRQLVDIRYKSDTRKFQKPLGAYGVKVELSYPDGSPAEG
VTVQIKAELTPKDNITYTSEVVSQRGLVGFEPIS IPTSAQHVLWLETKVMALNGKPVGAQYL
PSYLSLGSWYSPSQCYLQLQPPSHPLQVGEEAYFSVKSTPCPNFTLYYEVAARGNIVLSG
QQPAHTTQQRSKRAAPALEKPIRLTHLSETEPPEAPEAEVDVCVTSLHLAVTPSMVPLGR
LLVFYVRENGEGVADSLQFAVETFFENQVSVTYSANETQPGVEVDLIRIRAARGSCVCVAA
VDKSVYLLRSGFRLTPAQVFMQLEEDYDVSDSGFVGSREDGPFWWAGLTAQRRRRSSVFPWP
WGITKDSGFAFTETGLVVMTDRVSLNHRQDGGLYTDEAVPAFQPHTGSLVAVAPSRHPP
TEKRRKRTFFPETWIWHCLNISDPSNGHRLTSVKVPDSTISVWGEAVALSTSQGLGAEPSL
LKTFFKPFVDFMLPALIIRGEQVKIPLSVNYMGTCAEVYMKLSVPKGIQFVGHPGKRHV
TKKMCVAPGEAEPIWVLSFSDLGLNNTAKALAYGDTNCCDRGRSSKHPEENHADRRVP
IGVDHVRRSVMVEAEGVPRAYTYSAFFCPSERVHISTPNKYEFQYVQRPLRLTRFDVAVR
AHNDARVALSSGGPQDTAGMIEIVLGGHQNTRSWISTSKMGEPVASAHTAKILSWDEFRTF
WISWRGLIQVGHGPEPSNESIVAWTLPRPPEVQFIFGTFSTGWGSMGEFRIWRKMEVDES
YSEFTLGVPHGAIPGSERATASIGDVMGPTLNLHNLRLPFPGCGEQNMIHFAFNPFV
LYKLQKTKQQLSPEVERETTDYLVQGYQRQLTYKRQDGSYSAFGERDASGSMWLTAFLVLS
FAQARSFIFVDPRELAAKSWIIQQQQADGSFLAVGRVLNKDIQGGIHGIVPLTAYVVVA
LLETGTASEEERGSTDKARHFLESAAPLAMDYPSCALTTYALTLLRSPAAPALRKLRSI
AIMRDGVTHWSLSNSWDVDKGTPLSFSDRVSQSVVSAEVEMTAYALLTYTLLGDVAAALP
VVKWLSQQRNALGGSFSTQDTCVALQALAEYAILS YAGGINLTVSLASTNLDYQETFELH
RTNQKVLQATAIPLSTGLFVSAKGDGCCLMQIDVTYNVDPVAKPAQLLVLSLQPEAQ
GRPPMPASAAEGSRGDWPPADDDPAADQHHQEYKVMLEVCTRWLHAGSSNMMAVLEVP
LSGFRADIESLEQLLLLDKHMGMKRYEVAGRRVLFYFDEIPSRCLTCVRFRALRECVVGR
TALPVSVYDYYPFAFEATRFYNVSTHSPALRELCAAGPACNEVERAPARGPGWFPGESGPA
VAPEEGAAIARC GDHDCGAQGNPVCSDGVVYASACRLREACRQAAPLEPAPPSCCAL
EQRLPASSSSSTYGDDLASVAPGLQQDVKLNGAGLEVEDSDPEPEGEAEDRVTAGPRPPV
SSGNLESSTQSASPFHRWGQTPAPQRHSGRVVGAHRPGLLSPVFVYSPAFQSGGEEGLWM
SNTCTLR

A search of sequence databases reveals that the NOV11a amino acid sequence has 1794 of 1797 amino acid residues (99%) identical to, and 1796 of 1797 amino acid residues (99%) similar to, the 1884 amino acid residue ptnr:SPTREMBL-ACC:Q9ULD7 protein from Homo sapiens (Human) (KIAA1283 PROTEIN). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV11a is expressed in at least Adrenal Gland/Suprarenal gland, Bone Marrow, Brain, Heart, Kidney, Lung, Lymphoid tissue, Mammary gland/Breast, Pituitary Gland, Placenta, Prostate, Retina, Salivary Glands, Spleen, Thalamus, Thyroid. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

NOV 11b

A disclosed NOV11b nucleic acid of 6069 nucleotides (also referred to as CG57488_02) encoding a alpha-macroglobulin-like protein is shown in Table 11C. Putative untranslated regions upstream and/or downstream from the coding region, if any, are underlined, and the start and stop codons are in bold letters.

Table 11C. NOV11b nucleotide sequence (SEQ ID NO:27).

ATGAGCGGCGCCCTGCTCTGGCCGTTGCTCCCGCTCCTGCTCCTGCTGCTGTGCGCGCGG
GACGGCGTGCGCGCCGCGCAGCCTCAGGCCCGGGTTACTTGATTGCAGCTCCCTCTGTT
TTTCGCGCGGGCGTGAGGAAGTCATCAGCGTGACCATCTTTAACTCTCCAAGGGAAGTC
ACGGTCCAGGCTCAGCTGGTGGCCAGGGTGAGCCGGTGGTGAGAGCCAGGAGCCATC
CTGGATAAAGGGACAATCAAACCTCAAGGTGCCACGGGCCTCCGGGGCCAAGCGCTTCTG
AAAGTGTGGGGCCGCGCTGGCAGGCGGAGGAGGGCCCCCTCTTTCACAACCAGACCTCG
GTGACCGTGACGGCCGGGGCGCTTCTGTATTTCATCCAGACGGACAAGCCTGTGTACAGA
CCCCAGCACCGAGTGCTCATAAGCATCTTCACCGTCTCTCAAATCTGAGGCCTGTCAAC
GAGAAGCTGGAAGCCTACATCCTGGACCCCCGAGGCTCTCGGATGATAGAGTCGAGACAC
TTAAAGCCGTTCTGCTGCGGCATCACCAACATGAGCTTCCCCTTGTCCGACCAGCCTGTG
TTGGGAGAATGGTTCATTTTGTGAAATGCAAGGCCACGCGTACAACAAGTCTTTTGAA
GTTTCAGAAGTATGTGTTGCCAAGTTTGAGCTTCTGATTGACCCGCCCGGTATATCCAA
GACCTGGACGCCTGTGAGACAGGCACTGTGCGGGCCAGGTATACCTTTGGGAAACCTGTG
GCTGGTGCCTTAATGATCAACATGACTGTTAATGGTGTAGGGTACTACAGCCACGAGGTG
GGACGCCCTGTCTCAGAACAACCAAGATCCTCGGCTCCCGGACTTCGACATCTGCGTG
AGGGACATGATCCAGCGGACGCTCCCTGAGCACTTCCGGGGCAGGGTCAGCATCTGGGCC
ATGGTGACCAAGTGTGACGGGAGCCAGCAGGTGCGGTTGATGACTCCACCCCGTGCAG
AGGCAGCTGGTGGACATCCGGTACTCCAAGGACACGAGGAAGCAGTTCAGCCGGGCGCTG
GCCTACGTGGGGAAGGTGGAGCTATCCTACCCGATGGCAGCCAGCTGAGGGGGTGACG
GTCCAGATTAAGGCAGAGCTGACACCAAGGATAACATCTACACCAAGTGAAGTTGTGTCC
CAGCGTGGAAGTGTGGGGTTTGAAATCCCCTCCATCCCCACGTGAGCCAGCACGTGTGG
CTGGAGACCAAGGTGATGGCACTGAACGGGAAGCCCGTGGGGGCTCAGTACCTGCCCAGC
TACCTTCCCTCGGCAGCTGGTACTCCCCCAGCCAGTGCTACCTGCAGCTGCGCCACCC
TCCACCCACTGCAGGTTGGGGAAGAAGCCTATTTTCTGTGAAGTCCACATGTCCCTGC
AACTTTACCCTGTACTACGAGGTGGCTGCACGGGGCAATATTGTGCTATCGGGCCAGCAG
CCTGCCCACACCACCAGCAGCAAGCAAGCGGGCGGCCCTGCCCTGGAGAAACCGATT
CGTTTAACACACCTTTCTGAGACAGAGCCCCCACCAGCCCCAGAAGCTGAGTTCGACGTG
TGTGTGACCTCTCTTCATCTGGCCGTGACCCCCAGCATGGTCCCCCTTGGTCGCCTGCTG
GTCTTCTACGTGAGGAGAATGGAGAAGGGGTCGCGGACAGCCTTCAGTTTGAGTTCGAG
ACCTTCTTCGAAAACAGGTTTTCAGTGACGTATTACGAAATGAGACCAACCTGGGGAG
GTTGTGACCTGCGGATCAGGGCTGCAAGGGGCAGCTGTGTGTGCGTCCCGCAGTTGAT
AAGAGTGTCTACCTGCTCAGGTCTGGGTTCCGGCTGACTCCTGCCAGGTTTTCAGGAA
CTGGAAGATTATGATGTTTCTGATTCTTTGGCGTGTCCAGGGAGGATGGTCTTTTGG
TGGGCTGGGCTGACGGCACAACGACGCCGGCGCTCCTCTGTCTTCCCGTGGCCTTGGGGC
ATCACCAAGGACTCTGGGTTTGCTTACCGAAACGGGACTGGTGGTGTGACCGACCGA
GTGAGCCTGAACCACCGGAGGAGGCTGGCTCTACACCGATGAGGCTGTCCCCGCTTTC
CAGCCCCACACAGGGAGCCTGGTGGCAGTGGCTCCTTCCAGGCACCCCCCAGAACAGAG
AAGAGAAAAAGGACTTCTTCCCCGAAACATGGATTGGCATTGTCTCAACATCAGTGAC
CCATCTGGTGAGGGGACACTCAGTGTGAAGTCCCGGACTCCATCACCAGCTGGGTGGGT
GAGGCCGTGGCCCTGTCCACCTCTCAGGGCTTAGGCATCGCCGAGCCCTCCCTGCTGAAG
ACCTTCAAGCCCTTCTTGTGGACTTCATGCTCCCCGCTCTCATCATCCGTGGGGAGCAG
GTCAAGATCCCGCTCAGTGTCTACAACATACATGGGCACCTGCGCTGAGGTGTACATGAAG
CTCTCGGTTCCCAAGGCATCCAGTTTGTGGGCATCCTGGCAAACGCCATGTGACCAAG
AAGATGTGTGTGGCCCCGGGGAGGCTGAGCCCATCTGGGTCGTTCTGTCTTCTCAGCGAC
CTGGGACTCAACAACATCAGGCCAAAGCCCTTGCTTACGGAGACAAATGTCTGCGCG
GATGGGAGGTCCAGCAAACACCCTGAGGAGAATCACGCCGACAGGAGGTCCCCATCGGG
GTGGATCACGTGAGGCGCAGTGTGATGGTTGAGGCGGAAGGAGTCCCCGGGCGTACACC
TACAGCGCATTCTTCTGTCCCAGTGAGAGAGTCCACATCTCCACCCCCAACAGTATGAG
TTCCAGTATGTGCAGCGGCCACTGCGCCTCACCCGCTTTGATGTGGCTGTGCGAGCTCAC

AATGATGCCCGTGTGGCCTTGTCTTCTGGGCCCCAGGACACAGCAGGCATGATCGAGATC
 GTCCTGGGGGGGCATCAGAACACCAAGGTCTGGATCTCCACCAGCAAGATGGGAGAGCCC
 GTGGCCAGTGCACACACGGCCAAGATCCTCTCCTGGGATGAATTCAGAACATTCTGGATC
 AGCTGGCGTGGTGGCCTTATCCAGGTTGGCCATGGTCCAGAGCCATCCAATGAGTCTGTC
 ATTGTGGCCTGGACCCCTCCCGAGGCCACCAGAGGTCCAGTTCATTGGCTTTTCCACCGGC
 TGGGGCTCCATGGGTGAATTCGAATCTGGAGGAAGATGGAGGTGGACGAGAGCTACAGC
 GAGGCCTTCACCCTGGGGTCCCACACGGCGCCATCCCTGGGTCTGAGCGAGCCACCGCC
 TCCATCATCGGGGACGTTCATGGGGCCAACCCTGAACCACCTCAACAACCTCCTGCGGCTG
 CCGTTTGGCTGTGGAGAGCAGAACATGATCCACTTTGCACCCAACGTCTTTGTCTTGAAG
 TATCTTCAGAAAACCCAGCAGCTCAGCCCTGAGGTGGAGAGAGAGACCACCGATCCTA
 GTACAAGGCTACCAGCGCCAGCTGACCTACAAGCGCCAGGATGGCTCCTACAGCGCGTTT
 GGGGAGCGGGACGCATCGGGGAGCATGTGGCTCACAGCCTTTGTCTGAAGTCTTTCGCA
 CAGGCTCGCAGCTTTATCTTCGTGGACCCCCGGGAGCTGGCTGCCGCAAGAGCTGGATC
 ATCCAGCAGCAGCAGCGCGATGGCTCCTTCTGGCCGTGGGAGGGTCTTGAACAAGGAC
 ATCCAGGGTGGGATCCACGGCATTGTCCCGCTGACAGCCTACGTGGTGGTTGCTCTCCTG
 GAAACAGGCACAGCCTCAGAGGAGGAGAGAGGCTCCACTGACAAAGCGAGGCACTTCCTG
 GAGTCTGCTGCGCCCCCTGGCCATGGACCCCTTATAGCTGTGCCCTGACTACCTACGCGCTG
 ACCCTGCTCCGAGCCCCGAGCCCCCTGAGGCACCTGCGCAAGCTCCGTAGCCTGGCCATC
 ATGCGAGATGGGGTCACCCACTGGAGCCTGTCAAATTCCTGGGACGTGGACAAGGGCACA
 TTCTTGAGCTTCAGTGACAGGGTCTCTCAGTCAGTGGTCTCGGCCGAGGTGGAAATGACA
 GCCTACGCCCCTTCTGACCTACACTCTGCTGGGTGACGTGGCTGCCGCCCTGCCTGTGGTG
 AAGTGGCTGTCCAGCAGCGAAATGCACTTGGGGGCTTCTCCTCCACTCAGGACACCTGC
 GTGGCTCTGCAGGCCCTGGCTGAATATGCCATCTTGTCTATGCTGGAGGCATCAACCTC
 ACTGTCTCCCTGGCCTCCACCAACCTGGACTACCAGGAAACCTTCGAGCTGCACAGGACC
 AACCAGAAGGTCTTGACAGCAGCAGCGATCCCCAGCCTCCCCACGGGGCTGTTTGTGAGT
 GCCAAGGGGGACGGCTGCTGCTGATGCAGATTGATGTACCTACAATGTGCTGACCCG
 GTGGCCAAGCCAGCTTTCAGCTGCTCGTAAGCCTCCAGGAGCCTGAGGCCAGGGACGC
 CCGCCCCCATGCCCTGCCCTCCGAGCTGAGGGTTCCTGAGGAGACTGGCCCCCAGCTGAC
 GATGATGACCCAGCGCCGATCAGCATACCAGGAATACAAGGTGATGCTGGAGGTGTGC
 ACCAGGTGGCTGCATGCAGGGTCTTCCAATATGGCTGTCTTGGAGGTGCCCCCTGCTGTCA
 GGCTTCCGGGCAGACATCGAGAGCCTGGAGCAGCTGCTCCTTGACAAGCACATGGGGATG
 AAGAGGTATGAAGTGGCTGGACGCCGAGTGTCTTCTACTTTGATGAGATCCCCAGCCGG
 TGCTTGACCGTGCCTGCGGTTCCGTGCTCTCCGGGAGTGCCTGGTGGGAGGACGTGCGCG
 CTGCCAGTCTCCGTGTACGACTACTACGAACCCGCTTCGAGGCCACTCGCTTCTACAAC
 GTCAGCACGCACAGCCCACTCGCCCGGAACGTGTGCGCCGACCCGCGTGCAACGAAGTG
 GAGCGCGCCCCCTGCCCGGGCCCCGGGCTGGTTCCCCGGCGAGTCGGGCCCTGCCGTGGCC
 CCTGAGGAGGGGGCGCGATCGCGCGATGCGGCTGCGACACGACTGCGCGCCCCAGGGG
 AACCCGGTGTGCGGCTCCGACGGGGTGGTCTACGCCAGCGCCTGCCGCTGCGGGAGGCC
 GCCTGCCGCCAGGCCGCGCCCCCTGGAGCCCGCGCCTCCAGCTGCTGCGCCCTCGAGCAG
 CGCTTGCCGGCCTCGTCTCTCCACCTACGGGGATGACCTGGCTTCTGTGCCCCGGGG
 CTTTACAGCAGGACGTGAAGCTGAATGGAGCCGGCCTTGAGGTGGAGGACTCAGACCTT
 GAGCCTGAAGGGGAGGCGGAGGACAGGGTCACAGCCGGGCTCGGCCCTCTGTGAGCAGC
 GGGAACTGGAAAGCAGCACCCAGAGCGCCAGCCCGTTCCACAGATGGGGCCAGACTCCG
 GCCCCCTCAGAGACATAGTGGCCGGGTGGTGGGGGCCACAGGCCAGGGCTTCTGAGCCCT
 GTCTTCGTCTACAGCCAGCCTTTTCAGAGTGGTGGGAGGAGGGTTTATGGATGTCAAAC
 ACCTGCACCTTGAGATAATCCTACAACCACATGCAGTTGTGGGACCGAGTTTGGTCTTG
 GGGACATTACATACCCACACACCCAGCTTGTGCTGTGGTTAATCTCAGAAAACCTTG
 GTAAATGATCACTCCAGGATATTGACACGAATACACGTTACTGATCTTACTCACATGTTT
 TGGGGTGACATGAATTTGTGTGTGCATGTGTGTGTGTGTGCATGTGTGTGTCCCGGGC
 ACCTGACACCCCCAGCCAGGGCTGCCCAAAGTTGGGCTGATCAGAGACATAGACCCAAT
 GAGGAGCCCAACAGTGGCCCTCCAACCCTCTGCCCTGCCCCATAGTTTCATGCCCCAGTG
 GTCTTTGAAACTGCCCTGTGCCACTCCCTGGAGTGAGCAGCGGTGTCTCTGTGTGTGT
 GTGTCTGTG

In a search of public sequence databases, the NOV11b nucleic acid sequence, located on chromosome 19 has 5815 of 5817 bases (99%) identical to a gb:GENBANK-ID:AB033109|acc:AB033109.1 mRNA from Homo sapiens (Homo sapiens mRNA for

KIAA1283 protein, partial cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV11b polypeptide (SEQ ID NO:28) encoded by SEQ ID NO:27 has 1885 amino acid residues and is presented in Table D using the one-letter amino acid code.

- 5 Signal P, Psort and/or Hydropathy results predict that NOV11b has a signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.4600. The most likely cleavage site for a NOV11b peptide is between amino acids 25 and 26.

Table 11D. Encoded NOV11b protein sequence (SEQ ID NO:28).

```
MSGALLWPLLPLLLLLLSARDGVRAAQPOAPGYLIAAPSVFRAGVEEVISVTIFNSPREV
TVQAQLVAQGEPPVQSQGAILDKGTIKLKVPTGLRGQALLKVWGRGWQAEEGPLFHNQTS
VTVDGRGASVFIQTDPVYRPQHRVLISIFTVSPNLRPVNEKLEAYILDPRGSRMIEWRH
LKPFFCCGITNMSFPLSDQPVLGWEFIFVEMQGHAYNKSFEVQKYVLPKFELLIDPPRYIQ
DLDACETGTVRARYTFGKPVAGALMINMTVNGVGYYSHEVGRPVLRRTTKILGSRDFDICV
RDMIPADVPEHFRGRVSIWAMVTSVDGSQQVAFDDSTPVQRQLVDIRYSKDTRKQFKPGL
AYVGKVELSYPDGSPAEGVTQIKAELTPKDNIYTSEVVSQRGLVGFEIPSIPTSAQHVVW
LETKVMALNGKPVGAQYLPYSYLSLGSWYSPSQCYLQLQPPSHPLQVGEEAYFSVKSTCPC
NFTLYYEVAARGNIVLSGQQPAHTTQQRSKRAAPALEKPIRLTHLSETEPPEAPEAEVDV
CVTSLHLAVTPSMVPLGRLLVFYVRENGEGVADSLQFAVETFFENQVSVTYSANETQPGGE
VVDLIRARAARGSCVCAAVDKSVYLLRSGFRLTPAQVFQLEEDYDVSDFSFGVSREDGPFW
WAGLTAQRRRRSSVFPWPWGITKDSGFAFTETGLVVMTRVSLNHRQDGGGLYTDPAVPAF
QPHTGSLVAVAPSRHPPRTEKRRKRTFFPETWIWHCLNISDPSEGGLTSVKVPDSITSWVG
EAVALSTSQGLGIAEPSLLKTFKPFVDFMLPALIIRGEQVKIPLSVYNMGTCAEVYMK
LSVPKGIQFVGHPGKRHVTKKMCVAPGEAEPIWVLSFSDLGLNNITAKALAYGDTNCCR
DGRSSKHPEENHARRVPIGVHDVRRSVMVEAEGVPRAYTYSAFFCPSERVHISTPNKYE
FQYVQRPLRLTRFDVAVRAHNDARVALSSGPQDTAGMIEIVLGGHQNTRSWISTSKMGEP
VASAHTAKILSWDEFRTFWISWRGGLIQVGHGPEPSNESVIVAWTLPRPPEVQFIGFSTG
WGSMEGFRIWRKMEVDESYSFAFTLGVPHGAI PGSERATASIIIGDVMGPTLNHLNLLRL
PFGCGEQNMIHFAPNVFVLKYLQKTQQLSPEVERETTDYLVQGYQRQLTYKRQDGSYSF
GERDASGSMWLTAFVLKSFAQARSFIFVDPRELAASKSWIIQQQADGSFLAVGRVLNKD
IQGGIHGIVPLTAYVVVALLETGTASEEERGSTDKARHFLESAAPLAMDPYSCALTYYAL
TLLRSPAAPALRKLRLSLAIMRDGVTHWSLSNSWDVDKGTFLSFSDRVSQSVVSAEEMT
AYALLTYTLLGDVAAALPVVKWLSQQRNALGGFSSTQDTCVALQALAEYAILSYAGGINL
TVSLASTNLDYQETFELHRTNQKVLQTAAIPSLPTGLFVSAKGDGCLMQIDVTYNVPDP
VAKPAFQLLVSLQEPQAQGRPPMPASAAEGSRGDWPPADDDDDPAADQHHQYKVMLEVC
TRWLHAGSSNMAVLEVPLLSGFRADIESLEQLLLDKHMGMKRYEVAGRRVLFFYFDEIPSR
CLTCVRFRLRECVRGRTSALPVSVYDYEPAPFAEATRFYNVSTHSPARELCAGPACNEV
ERAPARGPGWFPGESGPAVAPEEGAAIARCGCDHDCGAQGNPVCSDGVVYASACRLREA
ACRQAAPLEPAPPSCCALEQRLPASSSSTYGDDLASVAPGPLQQDVKLNAGLEVEDSDP
EPEGEAEDRVTAGPRPPVSSGNLESSTQSASPFHRWGQTPAPQRHSGRVVGAHRPGLLSP
VFVYSPAFQSGGEEGLWMSNTCTLR
```

- 10 A search of sequence databases reveals that the NOV11b amino acid sequence has 1882 of 1884 amino acid residues (99%) identical to, and 1883 of 1884 amino acid residues (99%) similar to, the 1884 amino acid residue ptnr:SPTREMBL-ACC:Q9ULD7 protein from Homo sapiens (Human) (KIAA1283 PROTEIN). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

- 15 NOV11b is expressed in at least Adrenal Gland/Suprarenal gland, Bone Marrow, Brain, Heart, Kidney, Lung, Lymphoid tissue, Mammary gland/Breast, Pituitary Gland, Placenta, Prostate, Retina, Salivary Glands, Spleen, Thalamus, Thyroid. This information was

derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

5 NOV 11c

A disclosed NOV11c nucleic acid of 6157 nucleotides (also referred to as CG57488_03) encoding a alpha-macroglobulin-like protein is shown in Table 11E. Putative untranslated regions upstream and/or downstream from the coding region, if any, are underlined, and the start and stop codons are in bold letters.

Table 11E. NOV11C nucleotide sequence (SEQ ID NO:29).

GTGAGTAAGTGAGGGGACGATCCCCGGAAGGGATCGGGGCGGGTCTGGGCTCCGGAGATGG
GCGGAGCAGGCGTCCCGGGAGGGTGCGCCAGGAGCGGGGCGAGCGGGGCGAGCGGGGCG
GTCCCGGAGACGAGGCGGGTCCGGGAGGGGGCTGGCCCGGGGCTGCCAGCTTGGCCG
GGCGCGGAGCGGGGCGCATGGCGCCGGGCGCACTGCGCGGGGGCTGCGAACAAAGGGCCC
CCGCGCGGCGGCGGAGGACGGCGCGCTCGGAGCCCTGGCCCTGGCCAGCCCTGGCCCGG
CCCCCTCCCCAGGCGCGGCCCCCCCAGGAGCCGAAAAATGAGCGGCGCCCTGCTCTGGC
CGTTGCTCCCGCTCCTGCTCTCTGTGCTGTGCGGCGGGGACGGCTGCGCGCCGCGCAGC
CTCAGGCCCGGGTTACTTGATTGCAGCTCCCTCTGTTTTTCGCGCGGGCGTGGAGGAAG
TCATCAGCGTGACCATCTTTAACTCTCAAGGGAAGTCACGGTCCAGGCTCAGCTGGTGG
CCCAGGGTGAGCCGGTGGTGCAGAGCCAGGGAGCCATCCTGGATAAAGGGACAATCAAAC
TCAAGGTGCCACGGGCCCTCCGGGGCCAAGCGCTTCTGAAAGTGTGGGGCCGCGGCTGGC
AGGCGGAGGAGGGGCCCCCTTTTCAACACGAGCTCGGTGACCGTGGAGCGGCGGGGCG
CTCTGTATTATCAGCAGCGGAACGCTGTGTACAGACCCGAGCACCGAGTGTCTATAA
GCATCTTACCCTCTCTCCAAATCTGAGGCTGTGTAACGAGAAGCTGGAAGCCTACATCC
TGGACCCCCGAGGCTCTCGGATGATAGAGTGAGAGACACTTAAAGCCGTTCTGTGCGGCA
TCACCAACATGAGCTTCCCCCTGTCCGACCAGCCTGTGTTGGGAGAATGGTTCAATTTTG
TTGAAATGCAAGGCCACGCGTACAACAAGTCTTTTGAAGTTCAGAAGTATGTGTTGCCCA
AGTTTGAGCTTCTGATTGACCCGCCCCCGGTATATCCAAGACCTGGACGCTGTGAGACAG
GCACTGTGCGGGCAGGTATACCTTTGGGAAACCTGTGGCTGGTGCCTTAACGATCAACA
TGACTGTTAATGGTGTAGGTACTACGACCACGAGGTGGAGGACCTGTCTCTCAGAACAA
CCAGATCTCTCGGCTCCAGGACTTCGACATCTGCGTGGAGGACGATGCCAGCGGACG
TCCCTGAGCACTTCCGGGGCAGGGTCAGCATCTGGGCCATGGTGACCAGTGTGGACGGGA
GCCAGCAGGTGCGGTTGATGACTCCACCCCGTGCAGAGGCAGCTGGTGGACATCCGGT
ACTCCAAGGACACGAGGAAGCAGTTCAAGCCGGGCTGGCCTACGTGGGGGAAGGTGGAGC
TATCCTACCCCGATGGCAGCCAGCTGAGGGGGTGACGGTCCAGATTAAGGCAGAGCTGA
CACCAAAGGATAACATCTACACCAGTGAAGTTGTGTCCAGCGTGGACTAGTGGGGTTTG
AAATCCCCCTCATCCCCAGCTCAGCCGACGACGTGTGGCTGGAGACCAAGGTGATGGC
TGAACGGGAAGCCCGTGGGGCTCAGTACCTGCGTAGCTACTCTCCCTCGGCAGCTGGT
ACTCCCCCAGCCAGTGTACTTCGAGCTGCAGCCACCCCTCCACCCACTGCAGGTTGGGG
AAGAAGCCTATTTTTCTGTGAAGTCCACATGTCCCTGCAACTTTACCTGTACTACGAG
TGGCTGCACGGGGCAATATTGTGCTATCGGGCCAGCAGCCTGCCACACCACCCAGCAGC
GAAGCAAGCGGGCGGCCCTGCCCTGGAGAAACCGATTCTGTTTAAACACACCTTTCTGAGA
CAGAGCCCCCACCAGCCCCAGAAGCTGAGGTCGACGTGTGTGTGACCTCTCTTCATCTGG
CCGTGACCCCCAGCATGGTCCCCCTTGGTCGCTGCTGGTCTTCTACGTCAAGGAGAATG
GAGAAGGGGTGCGCGACAGCCCTTCAAGTTGACGTGACGACTCTTCTCGAAAACCAAGTTT
CAGTGACGTATTTCAGCAAATGAGACCCAACTGGGGAGGTTGTGACCTGCGGATCAGGG
CTGCAAGGGGCGAGCTGTGTGTGCGTGCCTGCGCAGTTGATAAGAGTGTCTACCTGCTCAGGT
CTGGGTTTCCGGCTGACTCTGCCAGGTTTTCCAGGAACTGGAAGATTATGATGTTTCTG
ATTCTTTTGGCGTGTCCAGGGAGGATGGTCCTTTTTTGGTGGGCTGGGCTGACGGCACAAC
GACGCCGCGCTCTCTGTCTTCCCGTGGCCTTGGGGCATCACCAAGGACTCTGGGTTTG
CCTTCCACGGAACCGGATGGTGGTGATGACCCAGCTGAGCCTGAACACCGGCGAGG
ACGGTGGCCTCTACACCGATGAGGCTGTCCCGCTTTCAGGCCCCACACAGGGAGCCTGG

TGGCAGTGGCTCCTTCCAGGCACCCCCCAGAACAGAGAAGAGAAAAGGACTTCTTCC
 CCGAAACATGGATTGGCATTGTCTCAACATCAGTGACCCATCTGGTGAGGGGACACTCA
 GTGTGAAGGTCCCGACTCCATCACCAGCTGGGTGGGTGAGGCCGTGGCCCTGTCCACCT
 CTCAGGGCTTAGGCATCGCCGAGCCCTCCCTGCTGAAGACCTTCAAGCCCTTCTTCGTGG
 ACTTCATGCTCCCCGCTCTCATCATCCGTGGGGAGCAGGTCAAGATCCCGCTCAGTGTCT
 ACAACTACATGGGCACCTGCGCTGAGGTGTACATGAAGCTCTCGGTTCCCAAGGCATCC
 AGTTTGTGGGCATCTGGCAAACGCCATGTGACCAAGAAGATGTGTGTGGCCCCCGGGG
 AGGCTGAGCCCATCTGGGTCTGTCTGTCTTCAGCGACCTGGGACTCAACAACATCACGG
 CCAAAGCCCTTGCTTACGGAGACACAAATTGCTGCCGGGATGGGAGGTCCAGCAAAACC
 CTGAGGAGAATCACGCCGACAGGAGGTCCCCATCGGGGTGGATCACGTCAGGCGCAGTG
 TGATGGTTGAGGCGAAGGAGTCCCCGGGCGTACACCTACAGCGCATTCTTGTCCCA
 GTGAGAGAGTCCACATCTCCACCCCCAACAAAGTATGAGTTCCAGTATGTGCAGCGGCCAC
 TGGCCTCACCCGCTTTGATGTGGCTGTGCGAGCTCACAATGATGCCCGTGTGGCCTTGT
 CTTCTGGGCCCCAGGACACAGCAGGCATGATCGAGATCGTCTTGGGGGGGCATCAGAACA
 CCAGGTCATGGATCTCCACCAGCAAGATGGGAGAGCCCGTGGCCAGTGCACACACGGCCA
 AGATCCTCTCTGGGATGAATTGAGAATCTTGATCAGCTGGCGTGGTGGCCTTATCC
 AGGTTGGCCATGGTCCAGAGCCATCCAATGAGTCTGTCTTGTGGCCTGGACCTCCCGA
 GGCCACCAGAGGTCCAGTTTATTGGCTTTTCCACCGGCTGGGGCTCCATGGGTGAATTCC
 GAATCTGGAGGAAGATGGAGGTGGACGAGAGCTACAGCGAGGCTTACCCTGGGGGTCC
 CACACGGCGCCATCCCTGGGTCTGAGCGAGCCACCGCTCCATCATCGGGGACGTCATGG
 GGCCAACCTGAACCACCTCAACAACCTCTGCGGCTGCCGTTTGGCTGTGGAGAGCAGA
 ACATGATCCACTTTGCACCCAACGTCTTTGTCTTGAAGTATCTTCAGAAAACCCAGCAGC
 TCAGCCCTGAGGTGGAGAGAGAGACCACCGACTACCTAGTACAAGGCTACCAGCGCCAGC
 TGACCTACAAGCGCCAGGATGGCTCTACAGCGCGTTTGGGGAGCGGGACGCATCGGGGA
 GCATGTGGCTCACAGCCTTGTCTGAAGTCTTTCGCACAGGCTCGCAGCTTTATCTTCG
 TGGACCCCGGAGCTGGCTGCGGCCAAGAGCTGGATCATCCAGCAGCAGCAGCGATG
 GCTCCTTCTGGCCGTGGGCAGGGTCTTGAACAAGGACATCCAGGGTGGGATCCACGGCA
 TTGTCCCGCTGACAGCCTACGTGGTGGTGTCTCTCTGGAAACAGGCACAGCCTCAGAGG
 AGGAGAGAGGCTCCACTGACAAAGCGAGGCATCTCTGGAGTCTGCTGCGCCCTGGCCA
 TGGACCTTATAGCTGTGCCCTGACTACCTACGCGCTGACCCTGCTCCGCAGCCCGGCAG
 CCCCTGAGGCATGCGCAAGCTCCGTAGCCTGGCCATCATGCGAGATGGGGTCACCCACT
 GGAGCCTGTCAAATTCCTGGGACGTGGACAAGGGCACATTCTTGAGCTTCAGTGACAGGG
 TCTCTCAGTCAGTGGTCTCGGCCGAGGTGGAATGACAGCCTACGCCCTTCTGACCTACA
 CTCTGCTGGGTGACGTGGCTGCGGCCCTGCCTGTGGTGAAGTGGCTGTCCAGCAGCGAA
 ATGCACTTGGGGGCTTCTCTCCACTCAGGACACCTGCGTGGCTCTGCAGGCCCTGGCTG
 AATATGCCATCTTGTCTATGCTGGAGGCATCAACCTCACTGTCTCCCTGGCCTCCACCA
 ACCTGGACTACCAGGAAACCTTCGAGCTGCACAGGACCAACCAGAAGGTTCTGCAGACAG
 CAGCGATCCCCAGCCTCCCCACGGGGCTGTTTGTGAGTGCCAAGGGGGACGGCTGTGCC
 TGATGCAGATTGATGTACCTACAATGTGCTGACCCGGTGGCCAAGCCAGCTTTCAGC
 TGCTCGTAAGCCTCCAGGAGCCTGAGGCCAGGGACGCCCGCCCCCATGCCTGCCCTCCG
 CAGCTGAGGGTTCCCGAGGAGACTGGCCCCCAGCTGACGATGATGACCCAGCGCGGCATC
 AGCATCACCAGGAATACAAGGTGATGCTGGAGGTGTGCACCAGGTGGCTGCATGCAGGGT
 CTTCCAATATGGCTGTCTTGGAGGTGCCCCGTGTGTGAGGCTTCCGGGCAGACATCGAGA
 GCCTGGAGCAGCTGTCTCTTGAACAAGCATGGGGATGAAGAGGTATGAAGTGGCTGGAC
 GCGGAGTGCTCTTCTACTTTGATGAGATCCCCAGCCGGTGCCTGACGTGCGTGGCTTCC
 GTGCTCTCCGGGAGTGCCTGGTGGGCAGGACGTGCGCGCTGCCAGTCTCCGTGTACGACT
 ACTACGAACCCGCTTCGAGGCCACTCGCTTCTACAACGTACGACGCACAGCCCACTCG
 CCCGGGAACGTGTGCGCCGACCCGCGTGCAACGAAGTGGAGCGCGCCCCCTGCCCGGGCC
 CGGGCTGGTTCCCGGCGAGTGGGGCCCTGCCGTGGCCCCCTGAGGAGGGGGCGGCGATCG
 CGCGATGCGGCTGCGACCACGACTGCGGCGCCAGGGGAACCCGGTGTGCGGCTCCGACG
 GGGTGGTCTACGCCAGCGCTGCCGCTGCGGGAGGCGCCTGCCGCCAGGCCGCGCCCC
 TGGAGCCCGCGCTCCAGCTGCTGCGCCCTCGAGCAGCGGCTGCCGGCTCTGTCGTCTCT
 CCACCTACGGGGATGACCTGGCTTCTGTGGCCCCGGGGCCTTTACAGCAGGACGTGAAGC
 TGAATGGAGCCGGCCTTGAGGTGGAGGACTCAGACCTGAGCCTGAAGGGAGGCGGAGG
 ACAGGGTCACAGCGGGCCTCGGCCCTCTGTGTGAGCAGCGGGAACCTGGAAAGCAGACCC
 AGAGCGCCAGCCCGTTCCACAGATGGGGCCAGACTCCGGCCCCCTCAGAGACATAGTGGCC
 GGGTGGTGGGGGCCACAGGCCAGGGCTTCTGAGCCCTGTCTTCTGTCTACAGCCAGCCT
 TTCAGAGTGGTGGGGAGGAGGGTTTATGGATGTCAAACACCTGCACCTTGAGATAATCCT
 ACAACCACATGCAGTTGTGGGACCGCAGTTTGGTCTTGGGGACCATTCATACCCACACAC
 CCAGCTTGTGCTGTGGTTAAACATCTCAGAAAACCTCTGGTAAATGATCACTCCAGGATAT
 TGACACGAATACACGTTACTGATCTTACTACATGTT

In a search of public sequence databases, the NOV11C nucleic acid sequence, located on chromosome 19 has 332 of 513 (64%) identical to a GENBANK-

ID:GPIMSPB|acc:D84339.1 *Cavia porcellus* mRNA for murinoglobulin. Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV11C polypeptide (SEQ ID NO:30) encoded by SEQ ID NO:29 has 1979 amino acid residues and is presented in Table 11F using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV11C has no signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.6000.

Table 11F. Encoded NOV11C protein sequence (SEQ ID NO:30).

```
MGGAGVPGGCAQERGERGERGGPGDEAGPGRGLARGCPSLAGRGAGRMAGALRGCEQR
APGGGARTAAALGPWPWPSPGPAPSPGAAPPRSRKMSGALLWPLLPLLLLLLSARDGVRAA
QPQAPGYLIAAPSVFRAGVEEVISVTIFNSPREVTVQAQLVAQGEVVSQGAILDKGTI
KLKVP TGLRGQALLKVWGRGWQAEEGPLFHNQTSVTVDGRGASVFIQTDKPVYRPOHRVL
ISIFTVSPNLRPVNEKLEAYILDPRGSRMIEWRHLKPFCCGITNMSFPLSDQPVLGWFI
FVEMQGHAYNKSFEVQKYVLPKFELLIDPPRYIQDLDACETGTVRARTFTGKPVAGALTI
NMTVNGVGYYSHVEVGRPVLRTTKILGSQDFDICVRDMPADVPEHFRGRVSIWAMVTSVD
GSQQVAFDDSTPVQRQLVDIRYSKDKTRQFKPGLAYVGKVELSYPDGSPAEGVTVQIKAE
LTPKDNIYTSEVVSQRGLVGFEIPIPTSAQHVLWLETKVMALNGKPVGAQYLPYSYLSLGS
WYSPSQCYLQLQPPSHPLQVGEEAYFSVKSTCPCNFTLYYEVAARGNIVLSGQQPAHTTQ
QRSKRAAPALEKPIRLTHLSETEPPPAPEAEVDVCVTS LHLAVTPSMVPLGRLLVFYVRE
NGEGVADSLQFAVETFFENQVSVTYSANETQPGEVVDLRIARAAGSCVCVAADVKS VYLL
RSGFRLTPAQVFQELEDDYDVSDFSFGVSREDGPFWWAGLTAQRRRRSSVFPWPWGITKDSG
FAFTETGLVVM TDRVSLNHRQDGGLYTDEAVPAFQPHGTSLVAVAPSRHPPRTEKRKRTF
FPETWIWHCLNIDSPSGEGTLSVKVPDSITSWVGEAVALSTSQGLGIAEPSLLKTFKPF
VDFMLPALIIRGEQVKIPLSVNYMGTCAEVYMKLSVPKGIQFVGHPGKRHVTKKMCVAP
GEAEP IWWVLSFSDLGLNNITAKALAYGDTNCCRDGRSSKHPEENHADRRVPIGVDHVR
SVMVEAEGVPRA YTYSAFFCPSE RVHISTPNKYEFQYVQRPLRLTRFDVAVRAHNDARVA
LSSGPQDTAGMIEIVLGGHQNTRSWISTSKMGEPVASAHTAKILSWDEFRTFWISWRGGL
IQVGHGPEPSNESVIVAWTLPRPPEVQFIGFSTGWGSMGEFRIWRKMEVDES YSEFTLG
VPHGAIPGSE RATA SIIGDVMGPTLNHLNLLRLPFGCGEQNMIFAPNVFVLKYLQKTQ
QLSPEVERETTDYLVQGYQRQLTYKRQDGSYSAFGERDASGSMWLTAFLVLSFAQARSFI
FVDPRELA AAKSWI IQQQADGSLAVGRVLNKDIQGGIHGIVPLTAYVVVVALLETGTAS
EEERGSTD K ARHFLESAAPLAMPYSCALTTYALTLLRSPAAPALRKLRLSLAIMRDGVT
HWSLSNSWDVDKGTFLSFSDRVSQSVVSAEVE MTAYALLTYTLLGDVAAALPVVKWLSQQ
RNALGGFSSTQDTCVALQALAEYAILSYAGGINLTVSLASTNLDYQETFELHRTNQKVLQ
TAAIPSLPTGLFVSAKGDGCCLMQIDVTYNVPDPVAKPAFQLLVSLQEPEAQGRPPMPA
SAAEGSRGDWPPADDDDPADQHHQEYKVMLEVCTRWLHAGSSNMAVLEVPLLSGFRADI
ESLEQLLLLDKHGMKRYEVAGRRVLFYFDEIPSRCLTCVRFRALREC VVGRTSALPVSVY
DYYPAPFAEATRFYNVSTHSP LARELCAGPACNEVERAPARGPGWFPGESGPAVAPEEGAA
IARCGCDHDCGAQGNPVCSDGVVYASACRLREAACRQAAPLEPAPPSCCALEQRLPASS
SSTYGDDLASVAPGPLQQDVKLNGAGLEVEDSDPEPEGEAEDRVTAGPRPPVSSGNLESS
TQSASPFHRWGQTPAPQRHSGRVVGAHRPGLLSPVFVYSPAFQSGGEEGLWMSNTCTLR
```

A search of sequence databases reveals that the NOV11C amino acid sequence has 171 of 432 amino acid residues (39%) identical to, and 258 of 432 amino acid residues (58%) similar to, the guinea pig protein ptnr:SPTREMBL-ACC:Q60486 ALPHA-

MACROGLOBULIN PRECURSOR - *Cavia porcellus*. Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

5 The disclosed NOV11a polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 11G.

Table 11G. BLAST results for NOV11					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 6331358 dbj BAA86597.1 (AB033109)	KIAA1283 protein [Homo sapiens]	1884	1882/1926 (97%)	1883/1926 (97%)	0.0
gi 15302736 ref XP_050563.2 (XM_050563)	KIAA1283 protein [Homo sapiens]	1711	1710/1711 (99%)	1710/1711 (99%)	0.0
gi 18567969 ref XP_095282.1 (XM_095282)	Similar to KIAA1283 protein [Homo sapiens]	837	665/732 (90%)	667/732 (90%)	0.0
gi 13928544 dbj BAB47146.1 (AB050668)	complement component C3 [Branchiostoma belcheri]	1732	220/611 (36%)	313/611 (51%),	4e-92
gi 17975514 ref NP_523506.1 (NM_078782)	Thiolester containing protein II [Drosophila melanogaster]	1420	195/583 (33%)	312/583 (53%)	4e-85

10 Tables 11H-I lists the domain descriptions from DOMAIN analysis results against NOV11. This indicates that the NOV11 sequence has properties similar to those of other proteins known to contain this domain.

Table 11H. Domain Analysis of NOV11
<p>gnl Pfam pfam00207, A2M, Alpha-2-macroglobulin family. This family includes the C-terminal region of the alpha-2-macroglobulin family. CD-Length = 751 residues, Score = 330 bits (847), Expect = 3e-91</p>

Table 11I. Domain Analysis of NOV11
<p>gnl Smart smart00280, KAZAL, Kazal type serine protease inhibitors; Kazal type serine protease inhibitors and follistatin-like domains. CD-Length = 46 residues, Score = 52.0 bits (123), Expect = 3e-07</p>

NOV11 is a member of the alpha-macroglobulin family. Alpha-macroglobulin proteins are large extracellular glycoproteins that can bind to and often act as reservoirs of growth factors and extracellular enzymes (See Gonias et al., J Biol Chem 2000 Feb 25;275(8):5826-31). Decreased level of these proteins in serum is often a sign of tissue damage (See Ruaux et al., Res Vet Sci 1999 Aug;67(1):83-7; Levine et al., J Pediatr Gastroenterol Nutr 1989 Nov;9(4):517-20; Wiedermann et al., Neoplasma 1978;25(2):189-96). These proteins may also help defend the body against bacterial or parasitic infection (See Araujo-Jorge et al., Parasitol Res 1992;78(3):215-21).

The disclosed NOV11 nucleic acid of the invention encoding an alpha-macroglobulin-like protein includes the nucleic acid whose sequence is provided in Table 11A, 11C or 11E or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 11A, 11C or 11E while still encoding a protein that maintains its alpha-macroglobulin-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 1 percent of the bases may be so changed.

The disclosed NOV11 protein of the invention includes the alpha-macroglobulin-like protein whose sequence is provided in Table 11B, 11D or 11F. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 11B, 11D or 11F while still encoding a protein that maintains its alpha-macroglobulin-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 1 percent of the residues may be so changed.

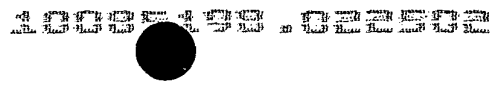
The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this alpha-macroglobulin-like protein (NOV11) may function as a member of a "alpha-macroglobulin

family". Therefore, the NOV11 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target
 5 (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV11 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the alpha-macroglobulin-like protein (NOV11) may be useful in gene therapy, and the alpha-macroglobulin-like protein (NOV11) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from cancer, trauma, regeneration (in vitro and in vivo),
 10 viral/bacterial/parasitic infections, adrenoleukodystrophy, congenital adrenal hyperplasia, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, graft versus host disease, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic
 15 stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, systemic lupus erythematosus, autoimmune disease, asthma, emphysema, scleroderma, allergy, ARDS, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, hypercalcaemia, Lesch-Nyhan syndrome, Von Hippel-Lindau (VHL) syndrome,
 20 diabetes, tuberous sclerosis, xerostomia, fertility, endocrine dysfunctions, growth and reproductive disorders, or other pathologies or conditions. The NOV11 nucleic acid encoding the alpha-macroglobulin-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV11 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV11 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV11 proteins have multiple hydrophilic
 30



regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

5 **NOV12**

NOV12 includes two orphan transporter-like proteins disclosed below. The disclosed sequences have been named NOV12a and NOV12b.

NOV12a

10 A disclosed NOV12a nucleic acid of 2119 nucleotides (also referred to as CG57526-01) encoding an orphan receptor-like protein is shown in Table 12A. Putative untranslated regions upstream and/or downstream from the coding region, if any, are underlined, and the start and stop codons are in bold letters.

Table 12A. NOV12a nucleotide sequence (SEQ ID NO:31).	
TGTGGTTTCCAAACGTCGGCAGAGGCTGGAGACGGCTCTCTAGTGCTGGGTGTGGAGTGA GGCACCCCTCGCCCTGAAGCCTGGGGCACTCAGTCACCATGGCTCATGCCCCAGAACC AGACCCGGCCGCGCAGCGACCTCGGGGATGAGAGGCCAAGTGGGACAACAAGGCCAGTA CCTCCTGAGCTGCATCGGGTTTGCCGTGGGGCTGGGAACATTTGGCGGTTCCCATACCT GTGCCAGACCTATGGAGGAGGTGCCTTCCTCATCCCCTACGTCATCGCGCTGGTCTTCGA GGGGATCCCCATTTCCACGTCGAGCTCGCCATCGGCCAGCGGCTGCGGAAGGGCAGCGT CGGCGTGTGGACGGCCATCTCCCCGTACCTCAGTGGAGTAGGTCTGGGCTGTGTACGCT GTCTTCTCTGATCAGCCTGTACTACAACACCATCGTGGCGTGGGTGCTGTGGTACCTCCT CAACTCCTTCCAGCACCCGCTGCCCTGGAGCTCCTGCCACCGGACCTCAACAGAACAGG TTTTGTGGAGGAGTGCCAGGGCAGCAGCGCCGTGAGCTACTTCTGGTACCGGCAGACACT GAACATCACAGCCGACATCAATGACAGTGGCTCCATCCAGTGGTGGCTGCTCATCTGCTT GGCAGCCTCCTGGGCAGTCGTGTACATGTGTGTCATCAGGGGCATTGAGACTACAGGGAA GGTGATTTACTTCACAGCTTTGTTCCCTTACCTGGTCTGACCATCTTCTCATCAGAGG GCTGACCCTGCCAGGGGCAACAAAAGGACTCATCTACTTGTTCACCTCCCAACATGCACAT TCTCCGAACCCCCGGGTGTGGCTGGACGCAGCCACCCAGATATTCTTCTCTGTGCCCT GGCCTTCGGAGGACACATCGCTTTTGCAAGTTACAACCTGCCAGGAGGAATGACTGCCA GAAGGATGCGGTGGTTCATCGCCCTGGTCAACAGGATGACCTCCCTGTACGCGTCCATCGC TGTCTTCTCTGCTCCTGGGGTTCAAAGCAACTAATGACCAGGAGCACTGCCTGGACAGGAA CATCCTCAGCCTCATCAACGACTTTGACTTCCCAGAGCAGAGCATCTCCAGGGACGACTA CCCAGCCGTCCTCATGCACCTGAACGCCACCTGGCCCCAAGAGGGTGGCCCAGCTCCCCCT GAAGGCCTGCCTCCTGGAAGACTTTCTGGATAAGAGTGCCTCGGGCCCGGGCCTGGCCTT CGTCGTCTTCACGGAGACCGACCTCCACATGCCGGGGGCTCCTGTGTGGGCCATGCTCTT CTTCGGGATGCTGTTCACCTTGGGGCTATCGACCATGTTTCGGGACCGTGGAGGCGGTTCAT CACACCCCTGCTGGACGTGGGGGTCTGCCTAGATGGGTCCCCAAGGAGGCCCTGACTGG TCCAGGGCTGGTCTGCCTGGTCTGCTTCTCTCCGCCACCTGCTTACGCTGCAGTCTGG GAACTACTGGCTGGAGATTTTCGACAATTTTGCCGCTTCCCTGAACCTGCTCATGTTGGC CTTTCTCGAGGTTGTGGGTGTGCTTATGTTTATGGAATGAAACGGTCTGCGATGACAT TGCGTGGATGACCGGGAGGCGGCCAGCCCCCTACTGGCGGCTGACCTGGAGGGTGGTCAG TCCCCTGCTGCTGACCATCTTTGTGGCTTACATCATCTCCTGTTCTGGAAGCCACTGAG ATACAAGGCTGGAACCCCAAATACGAGCTGTCCCCCTCGCGTCAGGAGAAGCTCTACCC GGGCTGGGCGCGCGCCGCTGTGTGCTGCTGTCTTGTGCTGCCGTGCTGTGGGTCCCGGT GGCCGCGCTTGTCTCAGTGTCTACCCGGCGGAGGCGGACGTGGAGGGACAGGGACGCGCG CCCAGACACGGACATGCGCCCGACACGGACACGCGCCAGACACGGACATGCGCCCGGA CACGGACATGCGCT TGA AGCCGGCGGAGCGGGCCTGCATGGGCGGGTCTGTGGGGGGC TTGGCCTGATGGTGGGCGGGGCCCCGCCACAGGGCCGACCCCAATACACCAGCGACTCA ACCTTAAAAAAAAAAAAA	

In a search of public sequence databases, the NOV12a nucleic acid sequence, located on chromosome 5 has 1122 of 1396 bases (80%) identical to a gb:GENBANK-ID:AF075263|acc:AF075263.1 mRNA from *Mus musculus* (*Mus musculus* orphan transporter isoform A11 (Xtrp2) mRNA, alternatively spliced, complete cds. Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV12a polypeptide (SEQ ID NO:32) encoded by SEQ ID NO:31 has 631 amino acid residues and is presented in Table 12B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV12a has a signal peptide and is likely to be localized to the plasma membrane with a certainty of 0.8000.

Table 12B. Encoded NOV12a protein sequence (SEQ ID NO:32).

MAHAPEPDPAASDLGDERPKWDNKAQYLLSCIGFAVGLGNIWRFPYLCQTYGGGAFLIPY
VIALVFEGIPIFHVELAIGQRLRKGSVGVWTAISPYLSGVGLGCVTLISFLISLYNTIVA
WVLWYLLNSFQHPLPWSSCPPDLNRTGFVEECQGSSAVSYFWYRQTLNITADINDSGSIQ
WWLLICLAASWAVVYMCVIRGIETTGVYFTALFPYLVLTIFLIRGLTLPGATKGLIYL
FTPNMHILQNPRVWLDAAATQIFFSLSLAFGGHIAFASYNPPRRNDCQKDAVVIALVNRMT
SLYASIAVFSVLGFKATNDQEHCLDRNLSLINDFDPEQSI SRDDYPVLMHLNATWPK
RVAQLPLKACLEDFLDKSASGPGLAFFVFTETDLHMPGAPVWAMLFFGMLFTLGLSTMF
GTVEAVITPLLDVGVLPWVPKEALTGPGLVCLVCFLSATCFTLQSGNYWLEIFDNFAAS
LNLLMLAFLEVGVVYVYGMKRFCDIAWMTGRRPSYWRLTWRVVSPLLLTIFVAYIIL
LFWKPLRYKAWNPKEYELFPSRQEKLYPGWARAACVLLSLLPVLWVPVAALAQLLTRRRRT
WRDRDARPDTPMRPDTPTRPDTPMRPDTPMR

A search of sequence databases reveals that the NOV12a amino acid sequence has 460 of 602 amino acid residues (76%) identical to, and 528 of 602 amino acid residues (87%) similar to, the 615 amino acid residue ptnr:SPTREMBL-ACC:O88576 protein from *Mus musculus* (Mouse) (ORPHAN TRANSPORTER ISOFORM A12). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV12a is expressed in at least Colon and Kidney. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

NOV12b

A disclosed NOV12b nucleic acid of 2039 nucleotides (also referred to as CG57526-02) encoding an orphan receptor-like protein is shown in Table 12C. Putative untranslated regions upstream and/or downstream from the coding region, if any, are underlined, and the start and stop codons are in bold letters.

Table 12C. NOV12b nucleotide sequence (SEQ ID NO:33).

TGTGGTTTCCAAACGTCGGCAGAGGCTGGAGACGGCTCTCTAGTGCTGGGTGTGGAGTGA
GGCACCACCCTCGCCCTGAAGCCTGGGGCACTCAGTCACCATGGCTCATGCCCCAGAACC
AGACCCGGCCCGCAGCGACCTCGGGGATGAGAGGCCAAGTGGGACAACAAGGCCAGTA
CCTCCTGAGCTGCATCGGGTTTCCCGTGGGGCTGGGGAACATTTGGCGGTTCACATACCT
GTGCCAGACCTATGGAGGAGGTGCCTTCCTCATCCCTACGTCATCGCGCTGGTCTTCGA
GGGGATCCCCATTTCCACGTCGAGCTCGCCATCGGCCAGCGGCTGCGGAAGGGCAGCGT
CGGCGTGTGGACGCCATCTCCCCGTACCTCAGTGGAGTAGGTCTGGGCTGTGTACGCT
GTCCTTCCTGATCAGCCTGTACTACAACACCATCGTGGCGTGGGTGCTGTGGTACCTCCT
CAACTCCTTCCAGACCCGCTGCCCTGGAGCTCCTGCCACCGGACCTCAACAGAACAGG
TTTTGTGGAGGAGTGCCAGGGCAGCAGCGCCGTGAGCTACTTCTGGTACCGGCAGACACT
GAACATCACAGCCGACATCAATGACAGTGGCTCCATCCAGTGGTGGCTGCTCATCTGCTT
GGCAGCCTCCTGGGCAGTCGTGTACATGTGTGTCATCAGGGGCATTGAGACTACAGGAA
GGTGATTTACTTACAGCTTTGTTCCCTTACCTGGTCTGACCATCTTCTCATCAGAGG
GCTGACCTGCCAGGGGCAACAAAAGGACTCATCTACTTGTTCCTCCCAACATGCACAT
TCTCCAGAACCCCGGGTGTGGCTGGACGCGAGCCACCCAGATATTCTTCTCTGTCCCT
GGCCTTCGGAGGACACATCGCTTTTGCAAGTTACAACCTCGCCAGGAGGAATGACTGCCA
GAAGGATGCGGTGGTTCATCGCCCTGGTCAACAGGATGACCTCCCTGTACGCGTCCATCGC
TGTCTTCTCTGTCTGGGGTTCAAAGCAACTAATGACCAGGAGCACTGCCTGGACAGGAA
CATCCTCAGCCTCATCAACGACTTTGACTTCCAGAGCAGAGCATCTCCAGGGACGACTA
CCCAGCCGTCCTCATGCACCTGAACGCCACCTGGCCCAAGAGGGTGGCCAGCTCCCCCT
GAAGGCTGCCCTCTGGAAGACTTTCTGGATAAGAGTGCCTCGGGCCCGGGCCTGGCCTT
CGTCGTCTTACGGAGACCGACCTCCACATGCCGGGGGCTCCTGTGTGGGCCATGCTCTT
CTTCGGGATGCTGTTACCTTGGGGCTATCGACCATGTTCCGGACCGTGGAGGCGGTCTAT
CACACCCTGCTGGACGTGGGGGTCTGCCTAGATGGGTCCCCAAGGAGGCCCTGACTGG
TCCAGGGCTGGTCTGCCTGGTCTGCTTCCCTCCTCCGCCACCTGCTTACGCTGCAGTCTGG
GAACCTACTGGCTGGAGATTTTCGACAATTTTGGCGCTTCCCTGAACCTGCTCATGTTGGC
CTTTCTCGAGGTGTGGGTGTGCTTTATGTTTATGGAATGAAACGGTTCTGCGATGACAT
TGCGTGGATGACCGGGAGCGGCCAGCCCCCTACTGGCGGTGACCTGGAGGGTGGTCTAG
TCCCCGTGCTGCTGACCATCTTGTGGCTTACATCATCTCCTGTTCTGGAAGCCACTGAG
ATACAAGGCCTGGAACCCCGAGGAGCTGTTCCCTCGCGTCAGGAGAAGCTCTACCCGGG
CTGGGCGCGCGCGCCTGTGTGCTGCTGCTTCTGCTGCCGTGCTGTGGGTCCCGGTGGC
CGCGCTTGCTCAGCTGCTACCCGGCGGAGGCGACGTGGAGACAGGCGCATGCTGAGGC
CGGGCTGGTGTTCAGGACTTCGAGAAGCAGAGGCCTGGCGTGGGGATACAGTACCTGAT
TCCAATGCTTTGCAACTTGCTCCAGACACTCTCCGGTAGAAAAAGAGCCTGTTCTTTT

In a search of public sequence databases, the NOV12b nucleic acid sequence, located on chromosome 5 has 1122 of 1396 bases (80%) identical to a gb:GENBANK-ID:AF075263|acc:AF075263.1 mRNA from *Mus musculus* (*Mus musculus* orphan transporter isoform A11 (Xtrp2) mRNA, alternatively spliced, complete cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV12b polypeptide (SEQ ID NO:34) encoded by SEQ ID NO:33 has 639 amino acid residues and is presented in Table 12D using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV12b has a signal peptide and is likely to be localized to the plasma membrane with a certainty of 0.8000.

Table 12D. Encoded NOV12b protein sequence (SEQ ID NO:34).

MAHAPEPDPAASDLGDERPKWDNKAQYLLSCIGFAVGLGNIWRFPYLCQTYGGGAFLIPY
VIALVFEGIPFIHVELAIGQRLRKGSVGVWTAISPYLSGVGLGCVTLISFLISLYNTIVA
WVLWYLLNSFQHPLPWSSCPPDLNRTGFVEECQSSAVSYFWYRQTLNITADINDSGSIQ
WWLLICLAASWAVVVMCIRGIETTGKVIYFTALFPYLVLTIFLIRGLTLPGATKGLIYL
FTPNMHILQNPRVWLDAAQIIFSLSLAFGGHIAFASYNSPRRNDCQKDAVVIALVNRMT
SLYASIAVFSVLGFKATNDQEHCLDRNLSLINDFDPEQSI SRDDYPAVLMHLNATWPK
RVAQLPLKACLEDFLDKSASGPGLAFVVFTETDLHMPGAPVWAMLFFGMLFTLGLSTMF
GTVEAVITPLLDVGVLPWPVPEALTGPGVLCLVCFLSATCFTLQSGNYWLEIFDNFAAS
LNLMLAFLEVGVVYVYGMKRFCDIAWMTGRRPSPYWRLTWRVVSPLLLTIFVAYIIL
LFWKPLRYKAWNPELFPQRQEKLYPGWARAACVLLSLPVLWVPVAALAQLLTRRRRTW

RQAHAEAGLVFQDFEKQRPVGVIQYLIPLCNLLQTLFR

A search of sequence databases reveals that the NOV12b amino acid sequence has 465 of 613 amino acid residues (75%) identical to, and 534 of 613 amino acid residues (87%) similar to, the 615 amino acid residue ptnr:SPTREMBL-ACC:O88576 protein from Mus musculus (Mouse) (ORPHAN TRANSPORTER ISOFORM A12)(. Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV12b is expressed in at least Colon and Kidney. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV12 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 12E.

Table 12E. BLAST results for NOV12					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 16550619 dbj BAB71018.1 (AK055798)	unnamed protein product [Homo sapiens]	628	626/631 (99%)	626/631 (99%)	0.0
gi 3347922 gb AAC27757.1 (AF075262)	orphan transporter isoform A12 [Mus musculus]	615	460/603 (76%)	528/603 (87%)	0.0
gi 8394204 ref NP_058859.1 (NM_017163)	X transporter protein 2 [Rattus norvegicus]	615	461/604 (76%)	528/604 (87%)	0.0
gi 3347924 gb AAC27758.1 (AF075263)	orphan transporter isoform A11 [Mus musculus]	605	454/603 (75%)	520/603 (85%)	0.0
gi 3347926 gb AAC27759.1 (AF075264) orphan	orphan transporter isoform B11 [Mus musculus]	577	416/603 (68%)	478/603 (78%)	0.0

Table 12G lists the domain descriptions from DOMAIN analysis results against NOV12. This indicates that the NOV12 sequence has properties similar to those of other proteins known to contain this domain.

Table 12G. Domain Analysis of NOV12

gnl|Pfam|pfam00209, SNF, Sodium:neurotransmitter symporter family.

CD-Length = 534 residues, 92.5% aligned

Score = 469 bits (1207), Expect = 2e-133

- A gene family encoding many Na(+)- and Cl(-)-dependent organic solute cotransporters has recently been recognized. Among the cotransporters that have been characterized are those for neurotransmitters, amino acids, and organic osmolytes. The cDNA
- 5 is 2,354 bp long with an open reading frame of 1,845 bp. The 615 deduced amino sequence shows ROSIT to be most clearly related to two orphan cDNAs of this family isolated from brain. Northern analysis showed the mRNA is normally expressed in renal cortex but not in brain, heart, colon, liver, stomach, or skeletal muscle. Moreover, hypernatremic rats displayed a marked increase in mRNA levels in renal cortex, renal outer medulla, and perhaps intestine.
- 10 Heterologous expression of the cRNA in *Xenopus laevis* oocytes failed to reveal the function of this gene product when analyzed with isotope fluxes or electrophysiological measurements using a wide variety of organic solutes. ROSIT is likely to be involved in kidney reclamation of an organic osmolyte or osmolyte precursor required for adaptation to hypertonic stress. (See Wasserman et al., *Am J Physiol* 1994 Oct;267(4 Pt 2):F688-94).
- 15 The disclosed NOV12 nucleic acid of the invention encoding a orphan receptor-like protein includes the nucleic acid whose sequence is provided in Table 12A or 12C or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 12A or 12C while still encoding a protein that maintains its orphan receptor-like activities and physiological
- 20 functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example,
- 25 modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 20 percent of the bases may be so changed.

The disclosed NOV12 protein of the invention includes the orphan receptor-like protein whose sequence is provided in Table 12B or 12D. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 12B or 12D while still encoding a protein that maintains its orphan
5 receptor-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 24 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this orphan receptor-like
10 protein (NOV12) may function as a member of a “orphan receptor family”. Therefore, the NOV12 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to:
15 protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV12 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies
20 and disorders as indicated below. For example, a cDNA encoding the orphan receptor-like protein (NOV12) may be useful in gene therapy, and the orphan receptor-like protein (NOV12) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from cancer, trauma, regeneration (in vitro and in vivo),
25 viral/bacterial/parasitic infections, Hirschsprung's disease, Crohn's Disease, appendicitis, diabetes, autoimmune disease, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, hypercalcaemia, Lesch-Nyhan syndrome, or other pathologies or conditions. The NOV12 nucleic acid encoding the orphan receptor-like protein of the invention, or fragments
30 thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV12 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV12 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods

5

NOV13

10

Table 13A. NOV13 nucleotide sequence (SEQ ID NO:35).

GTTACACCCCAAGACTAAGTTCTTTCCCAAGTTAGAGAAGAAGAGAGAAAGCAAAAAGAAG
 AGAGGAAAGTTCTCCCTTCCCTCCTCCGTGCCTGT**CAT**GTCTCTAAGCCAGAGCCGAA
 GGACGTCCACCAACTGAACGGGACTGGCCCTTCTGCCTCTCCCTGCTCTTCAGATGGCCC
 AGGGAGAGAGCCCTTGGCTGGGACCTCAGAGTTCTTGGGGCTGATGGGGCTGGGGTAGA
 GGTGGTGAATTGAGTCTCGGGCCAACGCCAAGGGGGTTTCGGGAGGAGGACGCCCTGCTGGGA
 GAACGGGAGCCAGAGCAACGAAGTGCAGCAGCTCAGCAGACAGCTGGCCCTCGCCAC
 TTCCCGCTCAAGGAGACCTCTTTCCATCGGGCTGCAAGTACTGTTTCCATTCTCTCT
 GGCAGGCTTTGGGACCGTGGCTGCTGGCATGGTGTGGACATCGTGCAGCACTGGGAAGT
 CTTCAGAAAGGTGACAGAGGTCTTCATCCTAGTGCCTGCGCTGCTGGGGCTCAAAGGGAA
 CCTGAAAATGACCTGGCATCAAGGCTTTCCACTGCAGCGAGTATCAACATTGGACACAT
 GGACACACCCAAGGAGCTCTGGCGGATGATCACTGGGAACATGGCCCTCATCCAGGTGCA
 GGCCACGGTGGTGGGCTTCTGGCGTCCATCGCAGCCGCTCGTCTTTGGCTGGATCCCTGA
 TGCCCACTTCAGTATTCCGACAGCCCTTCTGCTCTGTGCTAGCAGCTGGCCACAGCCCTT
 CATTGCTCCCTGGTACTGGGTATGATCATGATTGGAGTCATCATTTGGCTCTCGCAAGAT
 TGGATCAACCCAGACAATGTGGCCACACCCATTGCTGCCAGCCTGGGCGACCTCATCAC
 CTTGGCGCTGCTCTCAGGCATCAGCTGGGGACTCCTGACCTCTGCCCTCTCAGATCACTG
 GCGATACATCTACCCACTGGTGTGTGCTTTCTTTGTGGCCCTGCTGCCTGTCTGGGTGGT
 GCTGGCCCGACGAAGTCCAGCCACAAGGGAGGTGTTGTACTCGGGCTGGGAGCCTGTTAT
 CATTGCCATGGCCATCAGCAGTGTGGGAGGCCTCATCTTGGACAAGACTGTCTCAGACCC
 CAACTTTGCTGGGATGGCTGTCTTACAGCCTCTGTGATTAATGGTGTGGGGGCAATCTGGT
 GGCAGTGCAGCCAGCCGATCTCCACCTCTCTGCACATGAATGGAATGCCCGGAGAA
 CTTGAGCAAGCTCTCGCCGCTGTCCCACTTGTACCACCTTCTTCAGCCCTGGTGT
 GAATTCTCGCTCAGCCCGGGTCTCTTCTCTCTCGTGGTCCCAGGACACCTGGTGTTCCT
 CTACACCATCAGCTGTATGCAGGGCGGGCACACCACCTCACACTCATCTTCATCATCTT
 CTATATGACAGCTGCACTGCTCCAGGTGCTGATTCTCTGTACATCGCAGACTGGATGGT
 GCACTGGATGTGGGGCCGGGGCCTGGACCCGACAACCTTCTCCATCCCATACTTGACTGC
 TCTGGGGGACCTGCTTGGCACTGGGCTCCTAGCACTCAGCTTCCATGTTTCTCTGGCTCAT
 AGGGACCCGAGACACGGATGTGGGGACT**TAG**CTTGGTCACTCAACATTTTCCCCATCCCT
 CTGCACTTTCTATTTGAAATTTTCTTTTGTTCCTGTCCCTCCTCCACCCCACTCC
 CACCTCTT

15

fis, clone HEMBA1006926). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV13 polypeptide (SEQ ID NO:36) encoded by SEQ ID NO:35 has 517 amino acid residues and is presented in Table 13B using the one-letter amino acid code.

- 5 Signal P, Psort and/or Hydropathy results predict that NOV13 has a signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.6000.

Table 13B. Encoded NOV13 protein sequence (SEQ ID NO:36).

```
MSSKPEPKDVHQNLNGTGPSASPCSSDGPGRPLAGTSEFLGPDGAGVEVVIESRANAKGV
REEDALLENGSQSNESDDVSTDRGPAPPSPLKETSFSIGLQVLFPPFLLAGFGTVAAGMVL
DIVQHWEVFQKVTEVFILVPALLGLKGNLEMTLASRLSTAASINIGHMDTPKELWRMITG
NMALIQVQATVVGFLASIAAVVFGWIPDGHFSIPHAFLLCASSVATAFIASLVLMIMIG
VIIGSRKIGINPDNVATPIAASLGDLITLALLSGISWGLLTSALSDHWRYIYPLVCAFFV
ALLPVVWVVLARRSPATREVLVSGWEPVVIAMAISSVGGLLDKTVSDPNFAGMAVFTPTVI
NGVGGNLVAVQASRISTFLHMNGMPGENSEQAPRRCPSPCTTFFSPGVNSRSARVLFLLV
VPGHLVFLYTISCMQGGHTTLTLIFIIFYMTAALLQVLILLYIADWMVHWMWGRGLDPDN
FSIPYLTALGDLGTGLLALS FHVWLIGDRD TDVGD
```

- 10 A search of sequence databases reveals that the NOV13 amino acid sequence has 307 of 456 amino acid residues (67%) identical to, and 373 of 456 amino acid residues (81%) similar to, the 490 amino acid residue ptrn:TREMBLNEW-ACC:CAB66762 protein from Homo sapiens (Human) (HYPOTHETICAL 53.3 KDA PROTEIN). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

- 15 NOV13 is expressed in at least Liver, Pituitary Gland, Heart, Uterus, and B-cells. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV13 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 13C.

Table 13C. BLAST results for NOV13

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 14149819 ref NP_115524.1 (NM_032148)	hypothetical protein DKFZp434K0427 [Homo sapiens]	490	305/457 (66%)	372/457 (80%)	e-166
gi 12053165 emb CAB66762.1 (AL136828)	hypothetical protein [Homo sapiens]	490	305/457 (66%)	372/457 (80%),	e-166
gi 15079232 gb AAH11108.1 (BC011108)	protein for MGC:18986) [Mus musculus]	488	235/438 (53%)	306/438 (69%),	e-112

gi 14290540 gb AAH09039.1 AAH09039 (BC009039)	Similar to hypothetical protein FLJ20473 [Homo sapiens]	507	237/438 (54%)	306/438 (69%)	e-111
gi 12833620 dbj BAB22598.1 (AK003140)	homolog to CDNA FLJ12718 FIS, CLONE NT2RP1001286-putative [Mus musculus]	462	230/445 (51%)	304/445 (67%)	e-109

Table 13D lists the domain descriptions from DOMAIN analysis results against NOV13. This indicates that the NOV13 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 13D. Domain Analysis of NOV13
gnl Pfam pfam01769, MgtE, Divalent cation transporter. This region is the integral membrane part of the eubacterial MgtE family of magnesium transporters. Related regions are found also in archaeobacterial and eukaryotic proteins. All the archaeobacterial and eukaryotic examples have two copies of the region. This suggests that the eubacterial examples may act as dimers. Members of this family probably transport Mg ²⁺ or other divalent cations into the cell. The alignment contains two highly conserved Ds that may be involved in cation binding (Bateman A unpubl.)
CD-Length = 131 residues, 99.2% aligned
Score = 66.6 bits (161), Expect = 3e-12

A gene family encoding many Na(+)- and Cl(-)-dependent organic solute cotransporters has recently been recognized. Among the cotransporters that have been characterized are those for neurotransmitters, amino acids, and organic osmolytes. The cDNA is 2,354 bp long with an open reading frame of 1,845 bp. The 615 deduced amino sequence shows ROSIT to be most clearly related to two orphan cDNAs of this family isolated from brain. Northern analysis showed the mRNA is normally expressed in renal cortex but not in brain, heart, colon, liver, stomach, or skeletal muscle. Moreover, hypernatremic rats displayed a marked increase in mRNA levels in renal cortex, renal outer medulla, and perhaps intestine. Heterologous expression of the cRNA in *Xenopus laevis* oocytes failed to reveal the function of this gene product when analyzed with isotope fluxes or electrophysiological measurements using a wide variety of organic solutes. ROSIT is likely to be involved in kidney reclamation of an organic osmolyte or osmolyte precursor required for adaptation to hypertonic stress. (See Wasserman et al., Am J Physiol 1994 Oct;267(4 Pt 2):F688-94).

Nramp1 regulates macrophage activation in infectious and autoimmune diseases. Nramp2 controls anaemia. Both are divalent cation (Fe(2+), Zn(2+), and Mn(2+)) transporters;

Nramp2 a symporter of H(+) and metal ions, Nramp1 a H(+)/divalent cation antiporter. This provides a model for metal ion homeostasis in macrophages. Nramp2, localised to early endosomes, delivers extracellularly acquired divalent cations into the cytosol. Nramp1, localised to late endosomes/lysosomes, delivers divalent cations from the cytosol to phagolysosomes. Here, Fe(2+) generates antimicrobial hydroxyl radicals via the Fenton reaction. Zn(2+) and Mn(2+) may also influence endosomal metalloprotease activity and phagolysosome fusion. The many cellular functions dependent on metal ions as cofactors may explain the multiple pleiotropic effects of Nramp1, and its complex roles in infectious and autoimmune disease. (See Blackwell et al., *Microbes Infect* 2000 Mar;2(3):317-21).

Mutations in the gene encoding the renal epithelial K(+) channel ROMK1 (Kir 1.1) is one of the causes for Bartter's syndrome, an autosomal recessive disease. It results in defective renal tubular transport in the thick ascending limb of the loop of Henle that leads to hypokalemic metabolic alkalosis and loss of salt. Two novel ROMK1 mutations, L220F/A156V, have been described recently in a compound heterozygote patient demonstrating typical manifestations of Bartter's syndrome. Functional properties of these ROMK1 mutants were studied by coexpressing in *Xenopus* oocytes and by means of double electrode voltage clamp experiments. When both ROMK1 mutants were coexpressed no K(+) conductance could be detected. The same was found in oocytes expressing A156V-ROMK1 only or coexpressing wild type (wt) ROMK1 together with A156V-ROMK1. In contrast, K(+) conductances were indistinguishable from that of wt-ROMK1 when L220F-ROMK1 was expressed alone. Activation of protein kinase C signaling inhibited the conductance in both L220F-ROMK1 and wt-ROMK1 expressing oocytes. These effects were not seen in A156V-ROMK1 expressing oocytes.

The disclosed NOV13 nucleic acid of the invention encoding a cation transporter-like protein includes the nucleic acid whose sequence is provided in Table 13A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 13A while still encoding a protein that maintains its cation transporter-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or

derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 35 percent of the bases may be so changed.

5 The disclosed NOV13 protein of the invention includes the cation transporter-like protein whose sequence is provided in Table 13B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table B while still encoding a protein that maintains its cation transporter-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein,
10 up to about 33 percent of the residues may be so changed.

 The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

 The above defined information for this invention suggests that this cation transporter-like protein (NOV13) may function as a member of a "cation transporter family". Therefore,
15 the NOV13 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene
20 delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

 The NOV13 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the cation transporter-like
25 protein (NOV13) may be useful in gene therapy, and the cation transporter-like protein (NOV13) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from cancer, trauma, regeneration (in vitro and in vivo), viral/bacterial/parasitic infections, cardiomyopathy, atherosclerosis, hypertension, congenital
30 heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, transplantation, endometriosis, fertility, Von Hippel-Lindau (VHL) syndrome, cirrhosis, endocrine dysfunctions, diabetes, obesity, growth and reproductive disorders, or other pathologies or conditions. The NOV13

nucleic acid encoding the cation transporter-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV13 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV13 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV13 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV14

A disclosed NOV14 nucleic acid of 5175 nucleotides (also referred to as CG57593-01) encoding a ABC transporter-like protein is shown in Table 14A. Putative untranslated regions upstream and/or downstream from the coding region, if any, are underlined, and the start and stop codons are in bold letters.

Table 14A. NOV14 nucleotide sequence (SEQ ID NO:37).

AATGTTCAAAGGCTTTCTGTAAGTGAAGCTTTTTTTTTTCTTTTTTCCCCTAGCTATTGC
TCTGCAATATTTACTTTACCCCTGTTAATGAACAGGACAAAATGGTTAAAAAGAGATAAG
CGTGCGTCAACAAATTCAGGCTCTTCTGTACAAGAATTTCTTAAAAATGGAGAATAAA
AAGAGAGCTGGAGGAATGGACAATAACATTGTTTCTAGGGCTATATTTGTGCATCTTTTC
GGAACACTTCAGAGCTACCCGTTTTCTGAACAACCTCCTAAAGTCCTGGGAAGCGTGGA
TCAGTTTAATGACTCTGGCCTGGTAGTGGCATATACACCAGTCAGTAACATAACACAAAG
GATAATGAATAAGATGGCCTTGGCTTCTTTATGAAAGGTAGAACAGTCATTGGGACACC
AGATGAAGAGACCATGGATATAGAAGTCCAAAAAATACCATGAAATGGTGGGAGTTAT
ATTTAGTGATACTTTCTCATATCGCCTGAAGTTAATTGGGGATATAGAATCCCAGTTAT
AAAGGAGCACTCTGAATACACAGGTCACGTGTTGGGCCATGCATGGTGAAATTTTTTGTTA
CTTGGCAAAGTACTGGCTAAAAGGGTTTGTAGCTTTTCAAGCTGCAATTAATGCTGCAAT
TATAGAAGTAAGTACAACAAATCATTCTGTAATGGAGGAGTTGACATCAGTTATTGGAAT
AAATATGAAGATACCACCTTTCATTTCTAAGGGAGAAATTATGAATGAATGGTTTCATTT
TACTTGCTTAGTTTCTTCTCTTCTTTTATATACTTTGCATCATTTAAATGTTGCAAGGGA
AAGAGGAAAATTTAAGAACTGATGACAGTGATGGGTCTCCGAGAGTCAGCATTCTGGCT
CTCCTGGGGATTGACATACATTTGCTTCATCTTCATTATGTCCATTTTATGGCTCTGGT
CATAACATCAATCCCAATTGTATTTCACTGGCTTCATGGTGATATTCACACTCTATAG
CTTATATGGCCTTTCTTTGGTGTGGCTTTCTCATGAGTGTTTTAATAAGGAAACCTAT
GCTCGCTGGTTTGGCTGGATTTCTTCTCACTGTATTTTGGGGATGTCTGGGATCTCATGT
GTTATATAGACAACTTCCTTTATCTTTGGGATGGGTATTAAGTCTTCTTAGCCCTTTTGC
CTTCACTGCTGGAATACAGATTACACACCTGGATAATTACTTAAGTGGTGTTATTTTTCC
TGATCCCTCTGGGATTCATACAAATGATAGCCACTTTTTTCATTTGGCATTGATAC
TCTTTTCTATTTGATATTCACATTATATTTTGAGCGAGTTTTACCTGGTAAGGGCCATGG
GGATTCTCCATTATTTTCTTAAGTCCTCATTTTGGTCCAAACATCAAAATACTCATCA
TGAAATCTTTGAGAATGAAATAAATCCTGAGCATTCTCTGATGATTCTTTGAACCGGT
GTCTCCGAATTCATGGAAAAGAAGCCATAAGGATCAGAAATGTTATAAAAGAATATAA

In a search of public sequence databases, the NOV14 nucleic acid sequence, located on chromosome 17 has 1737 of 2520 bases (68%) identical to a gb:GENBANK-ID:AB020629|acc:AB020629.1 mRNA from Homo sapiens (Homo sapiens mRNA for KIAA0822 protein, complete cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV14 polypeptide (SEQ ID NO:38) encoded by SEQ ID NO:37 has 1595 amino acid residues and is presented in Table 14 using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV14 has a signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.8000. The most likely cleavage site for a NOV14 peptide is between amino acids 52 and 53.

Table 14B. Encoded NOV14 protein sequence (SEQ ID NO:38).

```
MVKKEISVRQQIQALLYKNFLKKWRIKRELEEWITITLFLGLYLICIFSEHFRATRFPPEQPP
KVLGSVDQFNDSGLVVAYTPVSNITQRIMNMALASFMKGRTVIGTPDEETMDIELPKKY
HEMVGVI FSDTFSYRLKFNWGYRIPVIKEHSEYTGHCWAMHGEIFCYLAKYWLKGFVAFQ
AAINAAII EVSTTNHSMVEELTSVIGINMKIPPFISKGEIMNEWFHFTCLVSFSSFIYFA
SLNVARER GKFKKLM TVMGLRESAFWLSWGLTYICFIFIMSIFMALVITSIPIVFHTGFM
VIFTLYSLYGLSLVLAFLMSVLIRKPMLAGLAGFLFTVFWGCLGFTVLYRQLPLSLGWVL
SLLSPFAFTAGIQITHLDNYLSGVIFPDPSGDSYKMIATFFILAFDTLFYLIPTLYFERV
LPGKGHGDSPLFFLKSSFWSKHQNTHEIFENEINPEHSSDDSFEPVSPPEFHGKEAIRIR
NVIKEYNGKTGKVEALQIFFDIYEGQITAILGHNGAGKSTLLNILSGLSVSTEGSATIYN
TQLSEITDMEEIRKNIGFCPOFNQFDFLTVRENLRVFAKIKGIQPKVEQEVL LLDDEPT
AGLDPF SRHRVWSLLKEHKVDRLILFSTQFMDEADILADRKVFLSNGKLCAGSSFLKR
KWGIGYHLSLHRNEMCDTEKITSLIKQHIPDAKLTTESEKLVYSLPLEKTNKFPDLYSD
LDKCS DQGIRNYAVSVTSLNEVFLNLEGKSAIDEPGIFDIGKQEKIHVTRNTGDESEMEQ
VLC SLPETRKAVSSAALWSRQIYAVATLRLFLKLRERRALLCLLLVLGIAFIP IILEKIM
YKVTRETHCWEFSPSMYFLSLEQIPKTPLTSLIVNNTGSGNIEDLVHSLKQDQIVLEIDD
FRNRNGSDDPSYNGAIIVSGDQKDYRFSVACNTKKSNCFPVLMGIVSNALIGIFNFTELI
QMESTFI FRDDIVLDLGFIDGSI FLLLITNCISPYIGITASVIIKVRGRERSQLWISGLW
PSAYWCGQALVDIPLYFLILFSIHLIYYFIFLGFQLSWELMFVLVSDPLFAGGMHNWLC S
PSYIPHICAFIHL SQVEKKNGFWSFGFFIVSYTCVHIYIKFYLLDKSFLPLVFTFNFYC
SYALMPVSCSKSVLLFAFLFLQKYLHIAPILFPLFAFVSVIFLFVIRCLEMKYGNEIMNKD
PVFRISPRSRETHPNPEEP EEEDEDVQAERVQAANALTAPNLEEEPVITASCLHKEYYET
KKSCFSTRKKKIAIRNVSFVKKGEVLGGLGHNGAGKSTS IKMITGCTKPTAGVVVVLQG
SRASVRQQHDNSLKF LGYCPQENSLWPKLT MKEHLELYAAVKGLGKEDAALSISRLVEAL
KLQEQLKAPVKTLSEGIKRKLCFVLSILGNPSVLLDEPFTGMDPEGQQQMWLQATVKNK
ERGTLLTTHYMSEAEAVCDRMAMMVSGTLRRCIGSIQHLKNKFGRDYLL EIKMKEPTQVE
ALHTEILKLF PQAAWQERYSSL MAYKLPVEDVHPLSRAFFKLERVKQTFNLEEYSLSQAT
LEQVFLELCKEQELGNVDDKIDTTVEWKLLPQEDP
```

A search of sequence databases reveals that the NOV14 amino acid sequence has 747 of 1321 amino acid residues (56%) identical to, and 951 of 1321 amino acid residues (71%) similar to, the 1581 amino acid residue ptnr:SPTREMBL-ACC:O94911 protein from Homo sapiens (Human) (KIAA0822 PROTEIN). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV14 is expressed in at least epidermis. This information was derived by determining the tissue sources of the sequences that were included in the invention including

Table 14E. Domain Analysis of NOV14

gnl|Smart|smart00382, AAA, ATPases associated with a variety of cellular activities; AAA - ATPases associated with a variety of cellular activities. This profile/alignment only detects a fraction of this vast family. The poorly conserved N-terminal helix is missing from the alignment.

CD-Length = 151 residues, 98.7% aligned

Score = 49.3 bits (116), Expect = 2e-06

ABCAI, a member of the ATP binding cassette family, mediates the efflux of excess cellular lipid to HDL and is defective in Tangier disease. The apolipoprotein acceptor specificity for lipid efflux by ABCAI was examined in stably transfected Hela cells, expressing a human ABCAI-GFP fusion protein. ApoA-I and all of the other exchangeable apolipoproteins tested (apoA-II, apoA-IV, apoC-I, apoC-II, apoC-III, apoE) showed greater than a threefold increase in cholesterol and phospholipid efflux from ABCAI-GFP transfected cells compared to control cells. Expression of ABCAI in Hela cells also resulted in a marked increase in specific binding of both apoA-I ($K_d = 0.60 \text{ ?g/mL}$) and apoA-II ($K_d = 0.58 \text{ ?g/mL}$) to a common binding site. In summary, ABCAI-mediated cellular binding of apolipoproteins and lipid efflux is not specific for only apoA-I but can also occur with other apolipoproteins that contain multiple amphipathic helical domains. (See Remaley et al., Biochem Biophys Res Commun 2001 Jan 26;280(3):818-823).

The molecular mechanisms regulating the amount of dietary cholesterol retained in the body, as well as the body's ability to exclude selectively other dietary sterols, are poorly understood. An average western diet will contain about 250-500 mg of dietary cholesterol and about 200-400 mg of non-cholesterol sterols. About 50-60% of the dietary cholesterol is absorbed and retained by the normal human body, but less than 1% of the non-cholesterol sterols are retained. Thus, there exists a subtle mechanism that allows the body to distinguish between cholesterol and non-cholesterol sterols. In sitosterolemia, a rare autosomal recessive disorder, affected individuals hyperabsorb not only cholesterol but also all other sterols, including plant and shellfish sterols from the intestine. The major plant sterol species is sitosterol; hence the name of the disorder. Consequently, patients with this disease have very high levels of plant sterols in the plasma and develop tendon and tuberous xanthomas, accelerated atherosclerosis, and premature coronary artery disease. The STSL locus was mapped to human chromosome 2p21 (ref. 4) and was localized it to a region of less than 2 cM bounded by markers D2S2294 and D2S2291. A new member of the ABC transporter family,

ABCG5, is mutant in nine unrelated sitosterolemia patients. (See Lee et al., Nat Genet 2001 Jan;27(1):79-83).

Pseudoxanthoma elasticum (PXE) is an inherited systemic disorder of connective tissue, characterized by progressive calcification of the elastic fibers in the eye, the skin, and the cardiovascular system. The PXE locus has been mapped to chromosome 16p13.1, and was recently further refined to a 500 kb-region, containing four candidate genes : MRP1 (ABCC1), MRP6 (ABCC6), pM5, and two copies of an unknown gene, the later subsequently found to be identical to the gene encoding the Nuclear Pore Interacting Protein (NPIP). In a comprehensive mutational screening, the entire coding region of the pM5, MRP1, and NPIP genes were analyzed in 7 patients affected with pseudoxanthoma elasticum, but failed to find evidence of disease-causing defects in any of these three genes. Five synonymous (G232G, P395P, A862A, G912G, D1106D), and five non synonymous (V404I, N458K, D490N, F1141I, G1195R) polymorphisms were found in the pM5 gene,

Mutations in the gene encoding ABCR (ABCA4), a photoreceptor-specific ATP-binding cassette (ABC) transporter, are responsible for autosomal recessive Stargardt disease (STGD), an early onset macular degeneration, and some forms of autosomal recessive cone-rod dystrophy and autosomal recessive retinitis pigmentosa. Heterozygosity for ABCA4 mutations may also represent a risk factor for age-related macular degeneration (AMD). An ongoing challenge in the analysis of ABCA4-based retinopathies arises from the observation that most of the ABCA4 sequence variants identified so far are missense mutations that are rare in both patient and control populations. With the current sample size of most sequence variants, one cannot determine statistically whether a particular sequence variant is pathogenic or neutral. A related challenge is to determine the degree to which each pathogenic variant impairs ABCR function, as genotype-phenotype analyses indicate that age of onset and disease severity correlate with different ABCA4 alleles. To address these questions, a functional analysis of human ABCR and its variants was performed. These experiments reveal a wide spectrum of biochemical defects in these variants and provide insight into the transport mechanism of ABCR. (See Sun et al., Nat Genet 2000 Oct;26(2):242-6).

The disclosed NOV14 nucleic acid of the invention encoding a ABC transporter-like protein includes the nucleic acid whose sequence is provided in Table 14A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 14A while still encoding a protein that maintains its ABC transporter-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are

complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 32 percent of the bases may be so changed.

The disclosed NOV14 protein of the invention includes the ABC transporter-like protein whose sequence is provided in Table 14B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 14B while still encoding a protein that maintains its ABC transporter-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 44 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this ABC transporter-like protein (NOV14) may function as a member of a "ABC transporter family". Therefore, the NOV14 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV14 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the ABC transporter-like protein (NOV14) may be useful in gene therapy, and the ABC transporter-like protein (NOV14) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from cancer, trauma, regeneration (in vitro and in vivo), viral/bacterial/parasitic infections, psoriasis, actinic keratosis, tuberous sclerosis, acne, hair

growth/loss, alopecia, pigmentation disorders, endocrine disorders, or other pathologies or conditions. The NOV14 nucleic acid encoding the ABC transporter-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

5 NOV14 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV14 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV14 proteins have multiple hydrophilic
10 regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV15

15 A disclosed NOV15 nucleic acid of 2540 nucleotides (also referred to as CG57652-01) encoding a diacylglycerol kinase alpha-like protein is shown in Table 15A. Putative untranslated regions upstream and/or downstream from the coding region, if any, are underlined, and the start and stop codons are in bold letters.

Table 15A. NOV nucleotide sequence (SEQ ID NO:39).

GGGGCGGTTCGACAGCTGAAGCAGGCCTACCCTCTGAAGAGGTCCAAGCAACGGAAGTACTA
CTACGAAGCTGCCTTTCTGGCCATCCTTGAGAAAAATAGACAGATGGCCAAGGAGAGGGG
CCTAATAAGCCCCAGTGATTTTGGCCAGCTGCAAAAATACATGGAATACTCCACCAAAAA
GGTCAGTGATGTCTAAAGCTCTTCGAGGATGGCGAGATGGCTAAATATGTCCAAGGAGA
TGCCATTGGGTACGAGGGATTCCAGCAATTCTGAAAATCTATCTCGAAGTGGATAATGT
TCCCAGACACCTAAGCCTGGCACTGTTTCAATCCTTTGAGACTGGTCACTGCTTAAATGA
GACAAATGTGACAAAAGATGTGGTGTGTCTCAATGATGTTTCTGCTACTTTTCCCTTCT
GGAGGGTGGTCGGCCAGAAGACAAGTTAGAATTACCTTCAAGCTGTACGACACGGACAG
AAATGGGATCCTGGACAGCTCAATGATGCGAGTGGCTGAATACCTGGATTGGGATGTGTC
TGAGCTGAGGCCGATTCTTCAGGAGATGATGAAAGAGATTGACTATGATGGCAGTGGCTC
TGTCTCTCAAGCTGAGTGGGTCCGGGCTGGGGCCACCACCGTGCCACTGCTAGTGCTGCT
GGGTCTGGAGATGACTCTGAAGGACGACGACAGCACATGTGGAGGCCCAAGAGGTTCCC
CAGACCAGTCTACTGCAATCTGTGCGAGTCAAGCATTGGTCTTGGCAAACAGGGACTGAG
CTGTAACCTCTGTAAGTACACTGTTACGACCAGTGTGCCATGAAAGCCCTGCCTTGTGA
AGTCAGCACCTATGCCAAGTCTCGAAGGACATTGGTGTCCAATCACATGTGTGGGTGCG
AGGAGGCTGTGAGTCCGGGCGCTGCGACCGCTGTGAGAAAAAGATCCGGATCTACCACAG
TCTGACCGGGCTGCATTGTGTATGGTGCCACCTAGAGATCCACGATGACTGCCTGCAAGC
GGTGGGCCATGAGTGTGACTGTGGGCTGCTCCGGGATCACATCCTGCCTCCATCTTCCAT
CTATCCCAGTGTCTGGCCTCTGGACCGGATCGTAAAAATAGCAAAACAAGCCAGAAGAC
CATGGATGATTTAAATTTGAGCACCTCTGAGGCTCTGCGGATTGACCTGTTCTTAACAC
CCACCCACTTCTCGTCTTTGTCAATCCTAAGAGTGGCGGGAAGCAGGGGCAGAGGGTGTCT
CTGGAAGTTCAGTATATATTAACCCCTCGACAGGTGTTCAACCTCCTAAAGGATGGTCC
TGAGATAGGGCTCCGATTATTCAAGGATGTTTCTGATAGCCGGATTTTGGTGTGTGGTGG
AGACGGCACAGTAGGCTGGATTCTAGAGACCATTGACAAAGCTAACTTGCCAGTTTTGGC

TCCTGTTGCTGTGTTGCCCTGGGTACTGGAATGATCTGGCTCGATGCCTAAGATGGGG
 AGGAGGTTATGAAGGACAGAATCTGGCAAAGATCCTCAAGGATTTAGAGATGAGTAAAGT
 GGTACATATGGATCGATGGTCTGTGGAGGTGATACCTCAACAACTGAAGAAAAAGTGA
 CCCAGTCCCCTTTCAAATCATCAATAACTACTTCTCTATTGGCGTGGATGCCTCTATTGC
 TCATCGATTCCACATCATGCGAGAGAAATATCCGGAGAAGTTCAACAGCAGAATGAAGAA
 CAAGCTATGGTACTTCAATTTGCCACATCTGAATCCATCTTCTCAACATGCAAAAAGCT
 GGAGGAGTCTTTGACAGTTGAGATCTGTGGGAAACCGCTGGATCTGAGCAACCTGTCCCT
 AGAAGGCATCGCAGTGCTAAACATCCCTAGCATGCATGGTGGCTCCAACCTCTGGGGTGA
 TACCAGGAGACCCCATGGGGATATCTATGGGATCAACCAGGCCCTAGGTGCTACAGCTAA
 AGTCATCACCGACCCTGATATCCTGAAAACCTGTGTACCAGACCTAAGTGACAAGAGACT
 GGAAGTGGTTGGGCTGGAGGGTGCAATTGAGATGGGCCAAATCTATACCAAGCTCAAGAA
 TGCTGGACGTCGGCTGGCCAAGTGCTCTGAGATCACCTTCCACACCACAAAAACCTTCC
 CATGCAAATTGACGTAGAACCCTGGATGCAGACGCCCTGTACAATCAAGATCACCCACAA
 GAACCAGATGCCCATGCTCATGGGGCCACCCCCCGCTCCACCAATTTCTTTGGCTTCTT
 GAGCTAAGGGGGACACCCTTGGCCTCCAAGCCAGCCTTGAACCCACCTCCCTGTCCCTGG
 ACTCTACTCCCAGGCTCTGTACATTGCTGCCACATACTCCTGCCAGCTTGGGGGAGTGT
 TCCTTCACCCTCACAGTATTTATTATCCTGCACCACCTCACTGTTCCCATGCGCACACA
 CATAACACACCCCAACACATACATTGAAAGTGCTCATCTGAATAAAATGACTTGTG
 TTTCCCTTTGGGATCTGCTG

In a search of public sequence databases, the NOV15 nucleic acid sequence, located on
 chromosome 12 has 2038 of 2038 bases (100%) identical to a gb:GENBANK-
 ID:HSDKRNA|acc:X62535.1 mRNA from Homo sapiens (H.sapiens mRNA for
 5 diacylglycerol kinase). Public nucleotide databases include all GenBank databases and the
 GeneSeq patent database.

The disclosed NOV15 polypeptide (SEQ ID NO:40) encoded by SEQ ID NO:39 has
 727 amino acid residues and is presented in Table 15B using the one-letter amino acid code.
 Signal P, Psort and/or Hydropathy results predict that NOV15 has no signal peptide and is
 10 likely to be localized in the nucleus with a certainty of 0.3000.

Table 15B. Encoded NOV15 protein sequence (SEQ ID NO:40).	
MAKERGLISPSDFAQLQKYMESTKKVSDVLKLFEDGEMAKYVQGDAGYEGFQQFLKIY	
LEVDNVPRHLSLALFQSFETGHCLNETNVTKDVCCLNDVSCYFSLLEGGRPEDKLEFTFK	
LYDTRNGILDSSMMRVAEYLDWDVSELRPILQEMMKEIDYDGSQSVSQAEWVRAGATTV	
PLLVLGLLEMTLKDDGQHMWRPKRFPVYCNLCESSIGLGKQGLSCNLCKYTVDQCAM	
KALPCEVSTYAKSRKDIGVQSHVWVRGGCESGRCDRCQKKIRIYHSLTGLHCVWCHLEIH	
DDCLQAVGHECDCGLLRDHILPPSSIYPSVLASGPRKNSKTSQKTMDDLNLSTSEALRI	
DPVPNTHPLLVFVNPKSGGKQQRVLWKFQYILNPRQVFNLLKDGPEIGLRLFKDVPDSR	
ILVCGDGTGVWILETIDKANLPVLPVAVLPLGTGNDLARCLRWGGGYEGQNLAKILKD	
LEMSKVHMDRWSVEVIPPQTEEKSDPVPFQIINNYFSIGVDASIAHRFHIMREKYPEKF	
NSRMKNKLWYFEFATSEISIFSTCKKLEESLTVEICGKPLDLNLSLEGIAVLNIPSMHGG	
SNLWGDTRRPHGDIYGINQALGATAKVITDPDILKTCVPDLSDKRLEVVGLEAIEMGQI	
YTKLKNAGRRLAKCSEITFHTTKTLPQIDVEPWMQTPCTIKITHKNQMPMLMGPPPRST	
NFFGFLS	

A search of sequence databases reveals that the NOV15 amino acid sequence has 727
 of 735 amino acid residues (98%) identical to, and 727 of 735 amino acid residues (98%)
 15 similar to, the 735 amino acid residue ptnr:SWISSNEW-ACC:P23743 protein from Homo
 sapiens (Human) (DIACYLGLYCEROL KINASE, ALPHA (EC 2.7.1.107) (DIGLYCERIDE

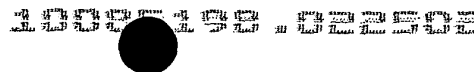
KINASE) (DGK- ALPHA) (DAG KINASE ALPHA) (80 KDA DIACYLGLYCEROL KINASE)). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV15 is expressed in at least Aorta, Appendix, Ascending Colon, Bone, Bone Marrow, Brain, Bronchus, Cartilage, Cervix, Colon, Coronary Artery, Dermis, Heart, Hippocampus, Kidney, Left cerebellum, Liver, Lung, Lymph node, Lymphoid tissue, Mammary gland/Breast, Ovary, Pancreas, Parotid Salivary glands, Peripheral Blood, Pituitary Gland, Placenta, Prostate, Respiratory Bronchiole, Small Intestine, Spleen, Stomach, Substantia Nigra, Synovium/Synovial membrane, Temporal Lobe, Testis, Thymus, Tonsils, Trachea, Umbilical Vein, Uterus, Vein, Whole Organism. Expression information was derived from the tissue sources of the sequences that were included in the derivation of the sequence of CG57652-01. The sequence is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:HSDKRNA|acc:X62535.1) a closely related H.sapiens mRNA for diacylglycerol kinase homolog in species Homo sapiens: lymphocytes, oligodendroglial cells, and neutrophils.. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV15 polypeptide has homology to the amino acid sequences shown
20 in the BLASTP data listed in Table 15C.

Table 15C. BLAST results for NOV15

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
<u>gi 11415024 ref NP 001336.1 </u> (NM_001345)	diacylglycerol kinase, alpha (80kD) [Homo sapiens]	735	727/735 (98%)	727/735 (98%)	0.0
<u>gi 125323 sp P20192 KDGA_PIG</u>	Diacylglycerol kinase, alpha (Diglyceride kinase) (DGK- alpha) (DAG kinase alpha) (80 kDa diacylglycerol kinase)	734	679/734 (92%)	703/734 (95%),	0.0
<u>gi 13879470 gb AAH0 6713.1 AAH06713</u> (BC006713)	diacylglycerol kinase, alpha (80 kDa) [Mus musculus]	730	597/736 (81%)	647/736 (87%),	0.0
<u>gi 18158459 ref NP 542965.1 </u> (NM_080787)	diacylglycerol kinase, alpha (80kD) [Rattus norvegicus]	727	600/736 (81%)	646/736 (87%),	0.0



gi 7949033 ref NP_058091.1 (NM_016811)	diacylglycerol kinase, alpha (80 kDa) [Mus musculus]	730	594/736 (80%)	645/736 (86%)	0.0
--	--	-----	---------------	---------------	-----

Tables 15D-E list the domain descriptions from DOMAIN analysis results against NOV15. This indicates that the NOV15 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 15D. Domain Analysis of NOV15

gnl|Pfam|pfam00609, DAGKa, Diacylglycerol kinase accessory domain (presumed). Diacylglycerol (DAG) is a second messenger that acts as a protein kinase C activator

CD-Length = 170 residues, 100.0% aligned

Score = 275 bits (702), Expect = 9e-75

Table 15E. Domain Analysis of NOV15

gnl|Smart|smart00109, C1, Protein kinase C conserved region 1 (C1) domains (Cysteine-rich domains); Some bind phorbol esters and diacylglycerol. Some bind RasGTP. Zinc-binding domains.

CD-Length = 50 residues, 96.0% aligned

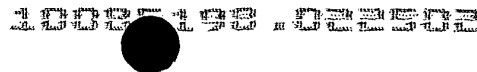
Score = 53.5 bits (127), Expect = 4e-08

Diacylglycerol (DAG) functions in intracellular signaling pathways as an allosteric
10 activator of protein kinase C (PKC; see 600448). In addition, DAG appears to play a role in
regulating RAS (see 190020) and RHO (see 165370) family proteins by activating the guanine
nucleotide exchange factors VAV (164875) and RASGRP1 (603962). DAG also occupies a
central position in the synthesis of major phospholipids and triacylglycerols. Thus, to maintain
cellular homeostasis, intracellular DAG levels must be tightly regulated (Topham and Prescott,
15 1999). DAG kinases (DGKs or DAGKs; EC 2.7.1.107) phosphorylate DAG to phosphatidic
acid, thus removing DAG. In intracellular signaling pathways, DAGK can be viewed as a
modulator that competes with PKC for the second messenger DAG. Schaap et al. (1990)
purified and characterized an 86-kD DAGK from normal human white blood cells. Based on
partial amino acid sequences of the purified enzyme, primers were designed that permitted
20 cloning of the human DAGK cDNA by use of PCR. The sequence demonstrated that it is the
human homolog of the porcine gene. The human DAGK cDNA, transfected into COS-7 cells,
resulted in a 6- to 7-fold increase in enzyme activity.

Several mammalian isozymes of DAGK have been identified. The isoform described by Schaap et al. (1990) has been designated DGK-alpha or DAGK1. Topham and Prescott (1999) stated that all DGKs have a conserved catalytic domain and at least 2 cysteine-rich regions homologous to the C1A and C1B motifs of PKCs. Most DGKs have structural motifs that are likely to play regulatory roles, and these motifs form the basis for dividing the DGKs into 5 subtypes. Type I DGKs, such as DGK-alpha, -beta (604070), and -gamma (601854), have calcium-binding EF-hand motifs at their N termini. DGK-delta (601826) and DKG-eta (604071) contain N-terminal pleckstrin homology (PH) domains and are defined as type II. DGK-epsilon (601440) contains no identifiable regulatory domains and is a type III DGK. The defining characteristic of type IV isozymes, such as DGK-zeta (601441) and -iota (604072), is that they have C-terminal ankyrin repeats. Group V is exemplified by DGK-theta (601207), which contains 3 cysteine-rich domains and a PH domain.

Pilz et al. (1995) pointed to the growing evidence to support some form of light-activated phosphoinositide signal transduction pathway in the mammalian retina. Although this pathway had no obvious role in mammalian phototransduction, mutations in this pathway were known to cause retinal degeneration in *Drosophila*. For example, the 'retinal degeneration A' mutant in *Drosophila* is caused by an alteration in the eye-specific DAGK gene. In an effort to consider genes mutated in *Drosophila* as candidates for mammalian eye disease, Pilz et al. (1995) determined the map position of 3 DAGK genes in the mouse. They localized the mouse homolog of DAGK1 to chromosome 10 by linkage analysis. By Southern blot analysis of human-hamster somatic cell hybrid DNA, Hart et al. (1994) assigned the DAGK gene to chromosome 12. Hart et al. (1994) further localized the gene to 12q13.3 by fluorescence in situ hybridization.

The disclosed NOV15 nucleic acid of the invention encoding a diacylglycerol kinase alpha-like protein includes the nucleic acid whose sequence is provided in Table 15A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 15A while still encoding a protein that maintains its diacylglycerol kinase alpha-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or



derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 0 percent of the bases may be so changed.

5 The disclosed NOV15 protein of the invention includes the diacylglycerol kinase alpha-like protein whose sequence is provided in Table 15B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 15B while still encoding a protein that maintains its diacylglycerol kinase alpha-like activities and physiological functions, or a functional fragment thereof. In
10 the mutant or variant protein, up to about 2 percent of the residues may be so changed.

 The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

 The above defined information for this invention suggests that this diacylglycerol kinase alpha-like protein (NOV15) may function as a member of a "diacylglycerol kinase
15 alpha family". Therefore, the NOV15 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic
20 marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

 The NOV15 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the diacylglycerol kinase
25 alpha-like protein (NOV15) may be useful in gene therapy, and the diacylglycerol kinase alpha-like protein (NOV15) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from osteoporosis, hypercalcaemia, arthritis, ankylosing
30 spondylitis, scoliosis, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, autoimmune disease, allergies, asthma, immunodeficiencies, transplantation, graft versus host disease, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neurodegeneration, tendonitis, fertility, atherosclerosis,

aneurysm, hypertension, fibromuscular dysplasia, scleroderma, myocardial infarction, embolism, cardiovascular disorders, bypass surgery, cardiomyopathy, atherosclerosis, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, renal tubular acidosis, IgA nephropathy, cirrhosis, systemic lupus erythematosus, emphysema, ARDS, lymphedema, endometriosis, diabetes, pancreatitis, obesity, anemia, ataxia-telangiectasia, endocrine dysfunctions, growth and reproductive disorders, inflammatory bowel disease, diverticular disease, ulcers, tonsillitis, ARDS, anemia, bleeding disorders, or other pathologies or conditions. The NOV15 nucleic acid encoding the diacylglycerol kinase alpha-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV15 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV15 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV15 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV16

A disclosed NOV16 nucleic acid of 3904 nucleotides (also referred to as CG57562-01) encoding a cation-transporting ATPase-like protein is shown in Table 16A. Putative untranslated regions upstream and/or downstream from the coding region, if any, are underlined, and the start and stop codons are in bold letters.

Table 16A. NOV16 nucleotide sequence (SEQ ID NO:41).

<p> <u>TTACCGGAAGTAA</u>AACTTCGGAAGTGAGGCGTTCCTCTGCCCGGAAGTGAGCGCGGCGCT AGGAAAGATGGCGGCAGCGGCGGCGGTGGGCAACGCGGTGCCCTGCGGGGCCCGGCCTTG CGGGGTCCGGCCTGACGGGCAGCCCAAGCCCGGGCGCAGCCGGCGCGCGCTCCTTGCCGC CGGGCCGGCGCTCATAGCGAACGGTGACGAGCTGGTGGCTGCCGTGTGGCCGTACGGCG GTTGGCGCTGTTGCGGCGCCTCACGGTGCTGCCATTGCGCGGGCTGCTTACCCGGCCTG GTTGGGTGCCGAGCCGCTGGCTGCTGGGGCTGGGGCAGCAGTTGGGTGCAGATCCCCGA AGCTGCGCTGCTCGTGCTTGCCACCATCTGCCTCGCGCACGCGCTCACTGTCTCTCGGG </p>

GCATTGGTCTGTGCACGCGCATTGCGCGCTCACCTGCACCCCGGAGTACGACCCAGCAA
 AGCGACCTTTGTGAAGGTGGTGCCAAACCCCAACAATGGCTCCACGGAGCTCGTGGCCCT
 GCACCGCAATGAGGGCGAAGACGGGCTTGAGGTGCTGCTTCGAATTCAGAAGATCAA
 GTATTCTACGATGCCCTGGAGAAGAAGCAGTTTCTCCCGTGCCCTTTCTGTGGGAAA
 CGCCTTCTCATACTATCAGAGCAACAGAGGCTTCCAGGAAGACTCAGAGATCCGAGCAGC
 TGAGAAGAAATTTGGGAGCAACAAGGCCGAGATGGTGGTGCCTGACTTCTCGGAGCTTTT
 CAAGGAGAGAGCCACAGCCCCCTTCTTTGTATTTCAGGTGTTCTGTGTGGGGCTCTGGTG
 CCTGGATGAGTACTGGTACTACAGCGTCTTTACGCTATCCATGCTGGTGGCGTTTCGAGGC
 CTCGCTGGTGCAGCAGCAGATGCGGAACATGTGCGAGATCCGGAAGATGGGCAACAAGCC
 CCACATGATCCAGGTCTACCGAAGCCGCAAGTGGAGGCCATTGCCAGTGATGAGATCGT
 ACCAGGGGACATCGTCTCCATCGGCCGCTCCCCACAGGAGAACCTGGTGCCATGTGACGT
 GCTTCTGCTGCGAGGCCGCTGCATCGTAGACGAGGCCATGCTCACGGGGAGTCCGTGCC
 ACAGATGAAGGAGCCCATCGAAGACCTCAGCCCAGACCGGGTGCTGGACCTCCAGGCTGA
 TTCCCGGCTGCACGTCTTTCGGGGGACCAAGGTGGTGCAGCACATCCCCCACAGAA
 AGCCACCACGGGCCCTGAAGCCGGTTGACAGCGGGTGCTGGCCTACGTCCTGCGGACCGG
 ATTCAACACATCCCAGGGCAAGCTGCTGCGCACCATCCTCTTTCGGGGTCAAGAGGGTGAC
 TGCGAACAACCTGGAGACCTTCATCTTCATCCTCTTCCTCCTGGTGTTCGCAATCGCTGC
 AGCTGCCTATGTATGGATTGAAGGTACCAAGGACCCAGCCGGAACCGCTACAAGATCGTT
 TCTGGAGTGCACCCTGATCCTCACCCTCGGTGCTGCCCTCTGAGCTGCCCATCGAGCTGTC
 CCTGGCCGTCAACACCTCCCTCATCGCCCTGGCCAAGCTCTACATGTACTGCACAGAGCC
 CTTCCGGATCCCCCTTTGCTGGCAAGGTGAGGTGTGCTGCTTTGACAAGACGGGGACGTT
 GACCAGTGACAGCCTGGTGGTGCAGCGGTGTGGCCGGGTGAGAGACGGGAAGGAGGTGAC
 CCCAGTGTCCAGCATCCCTGTAGAAACACACCGGGCCCTGGCCTCGTGCCACTCGCTCAT
 GCAGCTGGACGACGGCACCCCTCGTGGGTGACCTCTAGAGAAGGCCATGCTGACGGCCGT
 GGACTGGACGCTGACCAAGATGAGAAAGTATTCCCCGAAGTATTAAAACTCAGGGGCT
 GAAAATTCACCAGCGCTTTCATTTTGCCAGTGCCCTGAAGCGAATGTCCGTGCTTGCCTC
 GTATGAGAAGCTGGGCTCCACCGACCTCTGCTACATCGCGGCCGTGAAGGGGGCCCCGA
 AACTCTGCACTCCATGTTCTCCAGTGCCCGCCCGACTACCACCACATCCACACCGAGAT
 CTCCCGGGAAGGAGCCCGCTCCTGGCGCTGGGGTACAAGGAGCTGGGACACCTCACTCA
 CCAGCAGGCCCGGGAGGTCAAGCGGGAGGCCCTGGAGTGACGCTCAAGTTCGTGCGCTT
 CATTGTGGTCTCCTGCCCCGTCAAGGCTGACTCCAAGGCCGTGATCCGGGAGATCCAGAA
 TGCGTCCCACCGGGTGGTCATGATCACGGGAGACAACCCGCTCACTGCATGCCACGTGGC
 CCAGGAGTGCACCTTCATTGAAAAGGCCACACGCTGATCCTGCAGCCTCCCTCGAGAA
 AGGCCGGCAGTGCGAGTGGCGCTCCATTGACGGCAGCATCGTGCTGCCCCCTGGCCCGGGG
 CTCCCCAAAGGCACTGGCCCTGGAGTACGCACTGTGCCTCACAGGCGACGGCTTGGCCCA
 CCTGCAGGCCACCGACCCCAAGCAGCTGCTCCGCCTCATCCCCATGTGCAGGTGTTTCGC
 CCGTGTGGCTCCCAAGCAGAAGGAGTTTGTGATCACCAGCCTGAAGGAGCTGGGCTACGT
 GACCCTCATGTGTGGGATGGCACCAACGACGTGGGCCCTGAAGCATGCTGACGTGGG
 TGTGGCGCTCTTGGCCAATGCCCTGAGCGGGTTGTGAGCGGCGACGGCGGCCCGGGA
 CAGCCCAACCTGAGCAACAGTGCCATCAGAGCCACCTCCAGGACAGCCAAGCAGCGGT
 GGGGCTCCCTCCCTCCGAGGAGCAGCCAACCTCCAGAGGGACCGCTGAGCCAGGTGCT
 GCGAGACCTCGAGGACGAGAGTACGCCATTGTGAACTGGGGGATGCCAGCATCGCAGC
 ACCCTTACCTCCAAGCTCTCATCCATCCAGTGCATCTGCCACGTGATCAAGCAGGGCCG
 CTGCACGCTGGTGACCACGCTACAGATGTTCAAGATCCTGGCGCTCAATGCCCTCATCCT
 GGCCTACAGCCAGAGCGTCTCTACCTGGAGGGAGTCAAGTTTCAAGTACTTCCAGGCCAC
 CCTACAGGGGCTGCTGCTGGCCGGCTGCTTCTCTTCATCTCCCGTTCCAAGCCCCCAA
 GACCCTCTCCCGAGAACGGCCCCCTGCCCAACATCTTCAACCTGTACACCATCCTCACCGT
 CATGCTCCAGTTCTTTGTGCACTTCTTGAGCCTTGTCTACCTGTACCGTGAGGCCAGGC
 CCGGAGCCCCGAGAAGCAGGAGCAGTTCTGTGGACTTGTACAAGGAGTTTGAGCCAAGCCT
 GGTCAACAGCACCGTCTACATCATGGCCATGGCCATGCAGATGGCCACCTTCGCCATCAA
 TTACAAAGGCCCGCCCTTCATGGAGAGCCTGCCGAGAACAAGCCCCCTGGTGTGGAGTCT
 GGCAGTTTCACTCCTGGCCATCATTGGCCTGCTCCTCGGCTCCTCGCCCCACTTCAACAG
 CCAGTTTGGCCTCGTGACATCCCTGTGGAGTTCAAGCTGGTCAATTGCCAGGTCTTGCT
 CCTGGACTTCTGCTGGCGCTCCTGGCCGACCGCTCCTGCAGTTCTTCTGGGGACCCC
 GAAGCTGAAAGTGCTTCTTGGAGTGGTGTGGTACCCATGCCACCTGGCTGGCTGCC
 GCTGGGCGGGAACCCCAACAGGGCCCCGGGAGGGAACCCCTGCCCCCAACCCCCACAGCA
 AGGCTGTACAGTCTCGCCCTTGGAAAGACTGAGCTGGGACCCCCACAGCCATCCGCTGGCT
 TGGCCAGCAGAACCAGCCCCAAGCCAGCACCTTTGGTAAATAAAGCAGCATCTGAGATTT
 TAAA

In a search of public sequence databases, the NOV16 nucleic acid sequence, located on chromosome 19 has 3442 of 3442 bases (100%) identical to a gb:GENBANK-ID:AF288687|acc:AF288687.1 mRNA from Homo sapiens (Homo sapiens CGI-152 protein mRNA, complete cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV16 polypeptide (SEQ ID NO:42) encoded by SEQ ID NO:41 has 1204 amino acid residues and is presented in Table 16B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV16 has a signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.8000.

Table 16B. Encoded NOV16 protein sequence (SEQ ID NO:42).

```
MAAAAVGNAVPCGARPCGVRPDGQPKPGRSRRALLAAGPALIANGDELVAAVWPYRRLA
LLRRLTVLPFAGLLYPAWLGAAAAGCWGWGSSWVQIPEAALLVLATICLAHALTVLSGHW
SVHAHCALTCTPEYDPSKATFVKVPTPNNGSTELVALHRNEGEGLEVLSEFQKIKYS
YDALEKKQFLPVAFPVGNFASYYSNRGFGQEDSEIRAAEKKFGSNKAEMVVPDFSELFKE
RATAPFFVFQVFCVGLWCLDEYWYYSVFTLSMLVAFEASLVQQMRNMSEIRKMGKPHM
IQVYRSRKWRPIASDEIVPGDIVSIGRSPQENLVPCDVLRLRGRCIVDEAMLTGESVPQM
KEPIEDLSPDRVLDLQADSRLHVI FGGTKVVQHIPPQKATTGLKPVDSGCVAYVLRGTGN
TSQGKLLRTILFGVKRVTANNLETFFILFLLVFAIAAAAYVWIEGTKDPSNRNRYKLFLE
CTLILTSVVPPELPIELSLAVNTSLIALAKLYMYCTEPFRIPFAGKVEVCCFDKTGTLTS
DSLTVVRGVAGLRDGKEVTPVSSIPVETHRALASCHSLMQLDDGTLVGDPLEKAMLTAVDW
TLTKDEKVFPRSIKTQGLKIHQRHFHFAALKRMSVLASYEKLGSTDLCYIAAVKGAPETL
HSMFSQCPPDYHHIHTESISREGARVLALGYKELGHLTHQQAREVKREALECSLKFVGFIV
VSCPLKADSKAVIREIQNASHRVVMITGDNPLTACHVAQELHFIEKAHTLILQPPSEKGR
QCEWRSIDGSIVLPLARGSPKALALEYALCLTGDGLAHLQATDPQQLRLIPHVQVFARV
APKQKEFVITSLKELGYVTLMCGDGTNDVGALKHADVGVALLANAPERVVERRRRPRDSP
TLSNSGIRATSRTAKQRSGLPPEEQPTSQDRLSQVLRDLEDESTPIVKLGDAIAAPF
TSKLSSIQCICHVIKQGRCTLVTTLMFKILALNALILAYSQSVLYLEGVKFSDFOATLQ
GLLLAGCFLFISRSKPLKTLRERPLPNIFNLYTILTVMLQFFVHFLSLVLYLREAQARS
PEKQEQFVDLYKEFEPVLNSTVYIMAMAMQMATFAINYKGPFFMESLPENKPLVWSLAV
SLLAIIGLLLGSSPDFNSQFGLVDIPVEFKLVIAQVLLLDLDFCLALLADRVLFGLGTPKL
KVPS
```

A search of sequence databases reveals that the NOV16 amino acid sequence has 1141 of 1200 amino acid residues (95%) identical to, and 1164 of 1200 amino acid residues (97%) similar to, the 1200 amino acid residue ptnr:TREMBLNEW-ACC:BAB20095 protein from Mus musculus (Mouse) (CATION-TRANSPORTING ATPASE). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV16 is expressed in at least Adrenal Gland/Suprarenal gland, Amygdala, Aorta, Appendix, Artery, Bone, Bone Marrow, Brain, Bronchus, Brown adipose, Cartilage, Cerebral Medulla/Cerebral white matter, Cervix, Colon, Coronary Artery, Epidermis, Hair Follicles, Heart, Hippocampus, Kidney, Left cerebellum, Liver, Lung, Lymph node, Lymphoid tissue, Mammary gland/Breast, Ovary, Oviduct/Uterine Tube/Fallopian tube, Pancreas, Parietal Lobe, Peripheral Blood, Pituitary Gland, Placenta, Prostate, Respiratory Bronchiole, Right

Cerebellum, Skeletal Muscle, Skin, Spinal Cord, Spleen, Stomach, Substantia Nigra, Synovium/Synovial membrane, Temporal Lobe, Testis, Thymus, Urinary Bladder, Uterus, Vein, and Vulva. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV16 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 16C.

Table 16C. BLAST results for NOV16					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
<u>gi 14017867 dbj BAB47454.1 </u> (AB058728)	KIAA1825 protein [Homo sapiens]	1203	1200/1203 (99%)	1200/1203 (99%)	0.0
<u>gi 18202961 sp Q9HD20 </u> ATY2 HUMAN	Probable cation- transporting ATPase 2 (CGI- 152)	1204	1201/1204 (99%)	1201/1204 (99%)	0.0
<u>gi 18202867 sp Q9EP E9 </u> ATY2 MOUSE	Probable cation- transporting ATPase 2 (CATP)	1200	1141/1200 (95%)	1164/1200 (96%),	0.0
<u>gi 9966897 ref NP_0 65143.1 </u> (NM_020410)	CGI-152 protein [Homo sapiens]	1086	1072/1072 (100%)	1072/1072 (100%)	0.0
<u>gi 18467838 ref XP 079402.1 </u> (XM_079402)	BcdNA:GH06032 [Drosophila melanogaster]	1225	611/1125 (54%)	803/1125 (71%),	0.0

Tables 16D-E list the domain descriptions from DOMAIN analysis results against NOV16. This indicates that the NOV16 sequence has properties similar to those of other proteins known to contain this domain.

Table 16D. Domain Analysis of NOV16
<u>gnl Pfam pfam00122,</u> E1-E2_ATPase, E1-E2 ATPase CD-Length = 223 residues, 57.4% aligned Score = 64.3 bits (155), Expect = 4e-11

Table 16E. Domain Analysis of NOV16

[gnl|Pfam|pfam00702](#), Hydrolase, haloacid dehalogenase-like hydrolase. This family are structurally different from the alpha/ beta hydrolase family (pfam00561). This family includes L-2-haloacid dehalogenase, epoxide hydrolases and phosphatases. The structure of the family consists of two domains. One is an inserted four helix bundle, which is the least well conserved region of the alignment, between residues 16 and 96. The rest of the fold is composed of the core alpha/beta domain.

CD-Length = 197 residues, 78.2% aligned

Score = 39.7 bits (91), Expect = 0.001

Regulation of cation homeostasis is of critical importance to the cell and defects in proteins that regulate this process have been shown to cause a number of human diseases, including Darier-White disease, Menkes disease and Wilson disease. The human plasma

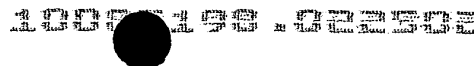
5 membrane $\text{Ca}(2+)$ -ATPase (PMCA) isoforms are members of the P class of ion-motive ATPases. PMCA removes bivalent calcium ions from eukaryotic cells and plays a critical role in intracellular calcium homeostasis by its capacity for removing calcium ions from cells against very large concentration gradients. Together with the highly related SERCA1 and SERCA3 isoforms encoded by ATP2A1 and ATP2A3, respectively, SERCA2 belongs to the

10 large family of P-type cation pumps that couple ATP hydrolysis with cation transport across membranes. SERCA pumps specifically maintain low cytosolic $\text{Ca}(2+)$ concentrations by actively transporting $\text{Ca}(2+)$ from the cytosol into the sarco/endoplasmic reticulum lumen. The ATP2A2 gene has been shown to be the site of mutations in Darier-White disease, an autosomal dominant skin disorder characterized by warty papules and plaques in seborrheic

15 areas (central trunk, flexures, scalp, and forehead), palmoplantar pits, and distinctive nail abnormalities (See Sakuntabhai et al., (1999) Mutations in ATP2A2, encoding a $\text{Ca}(2+)$ pump, cause Darier disease. Nature Genet. 21: 271-277). Patients with Menkes disease have mutations in the gene encoding $\text{Cu}(2+)$ -transporting ATPase alpha polypeptide and display early retardation in growth, peculiar hair, and focal cerebral and cerebellar degeneration ((See

20 Chelly et al., (1993) Isolation of a candidate gene for Menkes disease that encodes a potential heavy metal binding protein. Nature Genet. 3: 14-19). Wilson disease is an autosomal recessive disorder caused by mutations in the ATP7B gene, which encodes a copper-transporting ATPase, and is characterized by dramatic build-up of intracellular hepatic copper with subsequent hepatic and neurologic abnormalities (See Bull et al.. (1993) The Wilson

25 disease gene is a putative copper transporting P-type ATPase similar to the Menkes gene. Nature Genet. 5: 327-337).



The human cation transporting ATPase-like protein described in this invention is predicted to share the attributes of the other cation transporting ATPase family members and is thus implicated in the regulation of cation homeostasis. Given that a large number of cation transporting ATPases have been demonstrated to have a causative role in a variety of human diseases, the cation transporting ATPase-like protein is an attractive target for drug intervention in the treatment of human metabolic diseases, central nervous system disorders, immunological diseases and cancer, among others. The cation transporting ATPase-like gene described in this patent maps to human chromosome 19.

The disclosed NOV16 nucleic acid of the invention encoding a cation-transporting ATPase-like protein includes the nucleic acid whose sequence is provided in Table 16A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 16A while still encoding a protein that maintains its cation-transporting ATPase-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 0 percent of the bases may be so changed.

The disclosed NOV16 protein of the invention includes the cation-transporting ATPase-like protein whose sequence is provided in Table 16B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 16B while still encoding a protein that maintains its cation-transporting ATPase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 5 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this cation-transporting ATPase-like protein (NOV16) may function as a member of a "cation-transporting ATPase

family". Therefore, the NOV16 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target
 5 (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV16 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies
 10 and disorders as indicated below. For example, a cDNA encoding the cation-transporting ATPase-like protein (NOV16) may be useful in gene therapy, and the cation-transporting ATPase-like protein (NOV16) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from cancer, trauma, bacterial and viral infections, *in vitro*
 15 and *in vivo* regeneration, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, aneurysm, hypertension, fibromuscular dysplasia, stroke, obesity, transplantation, myocardial infarction, embolism, cardiovascular
 20 disorders, bypass surgery, anemia, bleeding disorders, adrenoleukodystrophy, congenital adrenal hyperplasia, diabetes, Von Hippel-Lindau (VHL) syndrome, pancreatitis, fertility, endometriosis, hypogonadism, hypercalcaemia, ulcers, cirrhosis, Hirschsprung's disease, Crohn's Disease, appendicitis, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, ataxia-telangiectasia, lymphedema, allergies, hemophilia, autoimmune disease,
 25 allergies, immunodeficiencies, transplantation, graft versus host disease (GVHD), osteoporosis, arthritis, ankylosing spondylitis, scoliosis, arthritis, tendonitis, muscular dystrophy, Lesch-Nyhan syndrome, myasthenia gravis, Alzheimer's disease, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, multiple sclerosis, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neurodegeneration, endocrine dysfunctions,
 30 growth and reproductive disorders, systemic lupus erythematosus, asthma, emphysema, ARDS, psoriasis, actinic keratosis, acne, hair growth/loss, alopecia, pigmentation disorders,, cystitis, incontinence, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, renal tubular acidosis, and IgA nephropathy, or other pathologies or conditions. The NOV16 nucleic acid encoding the cation-transporting ATPase-like protein of

the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV16 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV16 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV16 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV17

A disclosed NOV17 nucleic acid of 1167 nucleotides (also referred to as CG55914-01) encoding an acyl CoA desaturase-like protein is shown in Table 17A. Putative untranslated regions upstream and/or downstream from the coding region, if any, are underlined, and the start and stop codons are in bold letters.

Table 17A. NOV17 nucleotide sequence (SEQ ID NO:43).

ATGGGGCAGTTTTATAGGACTGGGGCAGGCAGTGGAAGTTACAGTTAAAGGTGGTTATC
TATTGTCAGCTGAGGAGGGATCACAAAGGTGAATGGTGAGGAGATCATAAGACTCATTGTC
CAAAAGAGGAATGTCACGAGGTCAATCGATCGATCAGTTGGGGCAGGGCAGGAACAAGTC
ATAATGGAATGTCTTAAGCCTTTTTACTTCAACTATCCATCTCTAGACAGTGAGGTCCTG
GATGATGACAGAGCCATAGATGGAAAAGACACCATTATTCTGGTCTATAAAGAACTGTCT
AGGGACTTGGCATCCTGTGTCCCAGCCACTCCAGCTGTGGCTGAAAGGGTCCAAGGTACA
GTTCAGGCCATGGCTTCAAAGGGTGCAAGCCCCAAGTCTCGGCAGCTTTCACAAGGTGTT
AAGCCTGGCAGTACAGAAGGGAAATGTGGGTTGGAGCCCCAACACAGAGCCCCACTGGGG
ACACTGCCTAGTGGAGCTTTGAGAAGAGGGCCACCATTCTCCAGACCCAGAATGGTAGA
CCCACTGACAGCTTGCACTATGCACTTGGAAAAGACACAGACACTCAACACCAGCCCATG
AAAGCAGCCAGAAGGGAGGCTGTACCTTGCAAGCTACAGGGGCAGAGCTGCCAAGACC
ATGGGAACCCAACCTGTTGCATCAGCATGACCCAGATGTGAGAATTGGAGTCAAAGAAGAT
CATTTTGGAGCTTTAAGATTTGACTGTCTTCTAGATTTTGGACTTACATGAGGACCCCA
GCCTTGCTGCTCTGCCATTGACCTCTGCCACCCTCCATACCGGGTGTGAGTTACCACCA
GAAGAAGTCTGTGGCAGCAGCAGTCTTTGCCTTATTCTGACAGTGCCAAAGTGTGTTTGG
CCCTACCAACAACCTGCAAGCCCTTTTCCTTGCTTTACTTTAGCTTTGGCTCCAGAAGGCTC
ATTGCTTTGAAATGTAACATATCCTGGTCCAACATACATCCGATTCCATGGGTCTGCATCA
TTATCTCCAAAACCTTCCAGTTGCATTGCTGTCATGTTTGTGTTAAAGAGACTGATGAGT
CTGGACTATGACCGGAAGAAAGTCTCCACGGCCCATCTTGCCAGGATTAAAGAAGCT
GGAGATGGAAGCTACAAGAGTGGCT**G**A

In a search of public sequence databases, the NOV17 nucleic acid sequence, located on chromosome 5 has 316 of 381 bases (82%) identical to a gb:GENBANK-
ID:AK000899|acc:AK000899.1 mRNA from Homo sapiens (Homo sapiens cDNA FLJ10037

Table 17C. BLAST results for NOV17

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 14756057 ref XP_052325.1 (XM_052325)	hypothetical protein FLJ14058 [Homo sapiens]	222	47/74 (63%)	53/74 (71%)	2e-18
gi 18572532 ref XP_088501.1 (XM_088501)	KIAA1161 protein [Homo sapiens]	188	59/145 (40%)	72/145 (48%)	7e-17
gi 14722528 ref XP_033352.1 (XM_033352)	similar to hypothetical protein FLJ14058 [Homo sapiens]	129	45/94(47%)	56/94 (58%)	2e-15
gi 16588389 gb AAL26787.1 AF304442.1 (AF304442)	B lymphocyte activation- related protein BC-1514 [Homo sapiens]	130	48/126 (38%)	56/126 (44%)	3e-10
gi 18562020 ref XP_087778.1 (XM_087778)	hypothetical protein XP_087778 [Homo sapiens]	216	36/68 (52%)	39/68 (56%)	2e-8

Fatty acid desaturases (ec 1.14.99.-) are enzymes that catalyze the insertion of a double bond at the delta position of fatty acids. There are two distinct families of fatty acid desaturases which do not seem to be evolutionary related.

5 Family 1 is composed of:

- Stearoyl-coa desaturase (scd) (ec 1.14.99.5). scd is a key regulatory enzyme of unsaturated fatty acid biosynthesis. scd introduces a cis double bond at the delta(9) position of fatty acyl-coa's such as palmitoleoyl- and oleoyl-coa. scd is a membrane-bound enzyme that is thought to function as a part of a multienzyme complex in the endoplasmic reticulum of
10 vertebrates and fungi.

Family 2 is composed of:

- Plants stearoyl-acyl-carrier-protein desaturase (ec 1.14.99.6), these enzymes catalyze the introduction of a double bond at the delta(9) position of stearoyl-acp to produce oleoyl-acp. this enzyme is responsible for the conversion of saturated fatty acids to unsaturated fatty
15 acids in the synthesis of vegetable oils. - cyanobacteria desa an enzyme that can introduce a second cis double bond at the delta(12) position of fatty acid bound to membranes glycerolipids. desa is involved in chilling tolerance; the phase transition temperature of lipids of cellular membranes being dependent on the degree of unsaturation of fatty acids of the membrane lipids.

The disclosed NOV17 nucleic acid of the invention encoding an acyl CoA desaturase-like protein includes the nucleic acid whose sequence is provided in Table 17A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 17A while still encoding a protein
5 that maintains its acyl CoA desaturase-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes
10 nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense
15 binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 18 percent of the bases may be so changed.

The disclosed NOV17 protein of the invention includes the acyl CoA desaturase-like protein whose sequence is provided in Table 17B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 17B while still encoding a protein that maintains its acyl CoA desaturase-like
20 activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 17 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this acyl CoA
25 desaturase-like protein (NOV17) may function as a member of a “acyl CoA desaturase family”. Therefore, the NOV17 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target
30 (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV17 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies

and disorders as indicated below. For example, a cDNA encoding the acyl CoA desaturase-like protein (NOV17) may be useful in gene therapy, and the acyl CoA desaturase-like protein (NOV17) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from Cardiomyopathy, Atherosclerosis, Hypertension, Congenital heart defects, Aortic stenosis, Atrial septal defect (ASD), Atrioventricular (A-V) canal defect, Ductus arteriosus, Pulmonary stenosis, Subaortic stenosis, Ventricular septal defect (VSD), valve diseases, Tuberous sclerosis, Scleroderma, Obesity, Transplantation, Von Hippel-Lindau (VHL) syndrome, Cirrhosis, Transplantation, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, or other pathologies or conditions. The NOV17 nucleic acid encoding the acyl CoA desaturase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV17 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV17 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV17 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV18

A disclosed NOV18 nucleic acid of 853 nucleotides (also referred to as CG57328-01) encoding a myo-inositol-1 (or 4) monophosphatase-like protein is shown in Table 18A. The start and stop codons are in bold letters.

Table 18A. NOV18 nucleotide sequence (SEQ ID NO:45).

ATGGCTGATCCTTGGCAGGAATGCATGGATTATGCAGTAACCCTAGCAAGACAAGCTGGA
GAGGTGGTTCATGATGCTCTTAAAAATGAAGTGAATGTTATCCTGAAAGGTTCTCCAGTT
GATTTGGTAACTGCTACTGACCAAAAAGTTGAAAAAATGCTTATCTCTTCCATAAAGGAA
AAGTATCCATCTCATAGGTATTTTTTTATTTGTGAGGAATCTGGCAGCTGGGGAAAAAGGT

GTCTTAAC TGACAACCCTACGTGGATCATTGACCCTATTGATGGAACAAC TAAGTTTGTC
 CATAGATTTCTTTTGTAGCTGTTTCGATTGGCCTTGTTGTAAATAAGAAGGTAGAATTT
 GGAGTTGTGTACAGTTGTGTGGAAGACAAGAGGTACACTGTCAGGAAAGGAAAAGGTGCC
 TTTTATAATGGTCAAAAAC TACAGGTTTCACAAGAAGGTGATATTACCAAATC ACTCTTG
 GTGACCGAGCTGGGCTATTGCAGAACATCAGAAATTGTAAGA ACTATTCTTTCCAATATG
 GAAAAGCTTTCTTGCAATTCCTATTTCACGGTATCCAGAGTGTTGGAACAGCAGCTACTAAT
 ATGTGCATTGCGGCAAGTGGAGGAGCAGAGGCATTTTATGAAATGGGAATTCAGTGTCTGG
 GATATTGCAGTAGCTGCCATTATTGTTACTGAAGCTGGTGGCGTGCTAATGGATGTTACT
 GGTGGACCATTCCATTTAATGTACGGAGAATAATTGCTGCAAATTGTACAGCATTAGCA
 GAAAGGATAGCCAAAGAAATTCAGGTAGCACCTTTTCAATGAGATGATGAAGATTAATTA
 CAGCAGCCTCATA

In a search of public sequence databases, the NOV18 nucleic acid sequence, located on
 chromosome 8 has 761 of 853 bases (89%) identical to a gb:GENBANK-
 ID:AF042729|acc:AF042729.2 mRNA from Homo sapiens (Homo sapiens lithium-sensitive
 5 myo-inositol monophosphatase A1 (IMPA1) mRNA, complete cds). Public nucleotide
 databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV18 polypeptide (SEQ ID NO:46) encoded by SEQ ID NO:45 has
 273 amino acid residues and is presented in Table 18B using the one-letter amino acid code.
 Signal P, Psort and/or Hydropathy results predict that NOV18 has no signal peptide and is
 10 likely to be localized in the cytoplasm with a certainty of 0.6500.

Table 18B. Encoded NOV18 protein sequence (SEQ ID NO:46).

MADPWQECMDYAVTLARQAGEVVHDA LKNEVNVI LKGS PVDLVTATDQKVEKMLISSIKE
 KYP SHRYFFIVRNLAAGEKGVLT DNPTWIIDPIDGTTKFVHRFPFVAVSIGLVVNKKVEF
 GVVYSCVEDKRYTVRKKGAFYNGQKLQVSQEGDITKSLLVTELG YCRTSEIVRTILSNM
 EKLSCIP IHGIQSVGTAATNMCIAASGGAEAFYEMGIHCWDIAVA AIIVTEAGGVLM DVT
 GGP FHLMSRRI I AANCTALAE RIAKEIQVAPFQ

A search of sequence databases reveals that the NOV18 amino acid sequence has 222
 of 273 amino acid residues (81%) identical to, and 240 of 273 amino acid residues (87%)
 similar to, the 277 amino acid residue ptnr:SWISSNEW-ACC:P29218 protein from Homo
 15 sapiens (Human) (MYO-INOSITOL-1(OR 4)-MONOPHOSPHATASE (EC 3.1.3.25)
 (IMPASE) (IMP) (INOSITOL MONOPHOSPHATASE) (LITHIUM-SENSITIVE MYO-
 INOSITOL MONOPHOSPHATASE A1)). Public amino acid databases include the GenBank
 databases, SwissProt, PDB and PIR.

NOV18 is expressed in at least Brain, Lung. This information was derived by
 20 determining the tissue sources of the sequences that were included in the invention including
 but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE
 sources.

The disclosed NOV18 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 18C.

Table 18C. BLAST results for NOV18					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 18570157 ref XP_095533.1 (XM_095533)	hypothetical protein XP_095533 [Homo sapiens]	512			
gi 5031789 ref NP_05527.1 (NM_005536)	inositol (myo)- 1 (or 4)- monophosphatase 1 [Homo sapiens]	277	222/273 (81%)	240/273 (87%),	e-124
gi 3914092 sp P97697 MYOP RAT	Myo-inositol-1 (or 4)- monophosphatase (IMPase) (IMP) (Inositol monophosphatase) (Lithium- sensitive myo- inositol monophosphatase A1)	277	222/273 (81%)	240/273 (87%),	e-124
gi 443382 pdb 2HHM A	Chain A, Human Inositol Monophosphatase (E.C.3.1.3.25) Dimer Complex With Gadolinium And Sulfate	276			
gi 127716 sp P20456 MYOP BOVIN	Myo-inositol-1 (or 4)- monophosphatase (IMPase) (IMP) (Inositol monophosphatase) (Lithium- sensitive myo- inositol monophosphatase A1)	277	209/273 (76%)	236/273 (85%),	e-119

5

Table 18D lists the domain descriptions from DOMAIN analysis results against NOV18. This indicates that the NOV18 sequence has properties similar to those of other proteins known to contain this domain.

Table 18D. Domain Analysis of NOV18	
gnl Pfam pfam00459 , inositol_P, Inositol monophosphatase family	
CD-Length = 270 residues, 97.0% aligned	
Score = 238 bits (607), Expect = 3e-64	

10

Myo-inositol-1(or 4)-monophosphatase enzyme catalyzes the reaction:

myo-inositol 1-monophosphate + H₂O \rightleftharpoons myo-inositol + phosphate · Acts on both enantiomers of myo-inositol-1-phosphate and myo- inositol 4-phosphate. It does not act on inositol bisphosphates, trisphosphates or tetrakisphosphates. It has been shown that several
 5 proteins share two sequence motifs. Two of these proteins are enzymes of the inositol phosphate second messenger signalling pathway:

- Vertebrate and plants inositol monophosphatase (EC 3.1.3.25). - Vertebrate inositol polyphosphate 1-phosphatase (EC 3.1.3.57).

Other proteins are:

10 - Bacterial protein cysQ. CysQ could help to control the pool of PAPS (3'-phosphoadenoside 5'-phosphosulfate), or be useful in sulfite synthesis. - Escherichia coli protein suhB. Mutations in suhB results in the enhanced synthesis of heat shock sigma factor (htpR). - Neurospora crassa protein Qa-X. Probably involved in quinate metabolism. - Emericella nidulans protein qutG. Probably involved in quinate metabolism. - Yeast
 15 HAL2/MET22 involved in salt tolerance as well as methionine biosynthesis. - Yeast hypothetical hypothetical protein YHR046c. - Caenorhabditis elegans hypothetical protein F13G3.5. - A Rhizobium leguminosarum hypothetical protein encoded upstream of the pss gene for exopolysaccharide synthesis. - Methanococcus jannaschii hypothetical protein MJ0109.

20 It is suggested that these proteins may act by enhancing the synthesis or degradation of phosphorylated messenger molecules. From the X-ray structure of human inositol monophosphatase, it seems that some of the conserved residues are involved in binding a metal ion and/or the phosphate group of the substrate.

The disclosed NOV18 nucleic acid of the invention encoding a myo-inositol-1 (or 4)
 25 monophosphatase-like protein includes the nucleic acid whose sequence is provided in Table 18A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 18A while still encoding a protein that maintains its myo-inositol-1 (or 4) monophosphatase-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes
 30 nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way

of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or
5 variant nucleic acids, and their complements, up to about 11 percent of the bases may be so changed.

The disclosed NOV18 protein of the invention includes the myo-inositol-1 (or 4) monophosphatase-like protein whose sequence is provided in Table 18B. The invention also includes a mutant or variant protein any of whose residues may be changed from the
10 corresponding residue shown in Table 18B while still encoding a protein that maintains its myo-inositol-1 (or 4) monophosphatase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 19 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or
15 (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this myo-inositol-1 (or 4) monophosphatase-like protein (NOV18) may function as a member of a “myo-inositol-1 (or 4) monophosphatase family”. Therefore, the NOV18 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various
20 pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to)
25 those defined here.

The NOV18 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the myo-inositol-1 (or 4) monophosphatase-like protein (NOV18) may be useful in gene therapy, and the myo-inositol-
30 1 (or 4) monophosphatase-like protein (NOV18) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from Systemic lupus erythematosus , Autoimmune disease, Asthma, Emphysema, Scleroderma, allergy, Von Hippel-Lindau (VHL) syndrome , Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's

disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, or other pathologies or conditions. The NOV18 nucleic acid encoding the myo-inositol-1 (or 4) monophosphatase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV18 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV18 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV18 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV19

A disclosed NOV19 nucleic acid of 2071 nucleotides (also referred to as CG57358-01) encoding a spinster-like protein is shown in Table 19A. The start and stop codons are in bold letters.

Table 19A. NOV19 nucleotide sequence (SEQ ID NO:47).

```

CCCCCGCGCGCCCGATCCGGGCGGGCATGATGTGCCTGGAATGCGCCTCGGCGGCGGCG
GGCGGCGCGGAGGAGGAGGAGGCGGACGCGGAGCGGCGGCGCGCGCGCGGGGGCGCAG
CGAGGGGCTGGCGGTAGCGGTTGCTGCGGGGCGCGGGGCGCGGGCGGCGCTGGAGTCTCG
GCCGCGGGCGATGAGGTGCAGACGCTGTGCGGCAGCGTAAGGCGGGCCCCGACCGGACCC
CCCGGCACCCCGGCACCCCGGCTGCGCAGCTACTGCAAAGGGCCCCGGCGCTCAGCAG
CCCAAACCGGCCAGCTTGGGCCGCGGGCGGGGCGAGCCGCCCATCTCAGCTTGGGC
AACGTGCTCAACTACCTGGACAGGTACACCGTGGCAGGCGTCCTTCTGGACATCCAGCAG
CACTTTGGGGTCAAGGACCGAGGCGCCGGCCTGCTGCAGTCAGTGTTTCATCTGTAGCTTC
ATGGTGGCTGCCCCATCTTCGGCTACCTGGGCGACCGCTTCAACAGGAAGGTGATTCTC
AGCTGCGGCATTTTCTTCTGGTCCGCCGTACCTTCTCCAGCTCCTTCATTCCCCAGCAG
TACTTCTGGCTGCTGGTCCCTGTCCCGGGGGCTGGTGGGCATCGGGGAGGCCAGCTACTCC
ACCATCGCCCCCACTATCATTTGGCGACCTCTTCACCAAGAACACGCGTACGCTCATGCTG
TCCGTCTTCTACTTCGCCATCCCACTGGGCAGTGGCCTGGGCTACATTACTGGCTCCAGC
GTGAAGCAGGCAGCCGAGACTGGCACTGGGCATTGCGGGTGTCCCTGTCTTGGGCATG
ATCACAGGAACACTCATCTCATTTCTGGTCCCAGCCACTAAAAGGGGTATGCCCGACCA
CTCGGGGACCAGCTCAAGGCCCGGACCTCATGGCTCCGAGATATGAAGGCCCTGATTCTGA
AACCGCAGCTACGTCTTCTCCTCCCTGGCCACGTCCGGCTGTCTCCTTCGCCACGGGGGCC
CTGGGCATGTGGATCCCGCTCTACCTGCACCGCGCCCAAGTTGTGCAGAAGACAGCAGAG
ACGTGCAACAGCCCGCCCTGTGGGGCCAAGGACAGCCTCATCTTGGGGCCATCACCTGC
TTTACGGGATTTCTGGGCGTGGTCACGGGGGCGAGGCCACGCGCTGGTGCCGCTGAAG
ACCCAGCGGGCCGACCACTGGTGTGTGCCGTGGGCATGCTGGGCTCTGCCATCTTCATC

```


TGCCTGATCTTCGTGGCTGCCAAGAGCAGCATCGTAGGAGCCTATATCTGTATCTTCGTC
 GGGGAGACGCTGCTGTTTTCTAACTGGGCCATCACTGCAGACATCCTCATGTACGTGGTC
 ATCCCCACGCGCGCGCCACTGCCGTGGCCTTGACAGAGCTTACCTCCCACCTGCTGGGG
 GACGCCGGGAGCCCCCTACCTCATTGGCTTTATCTCAGACCTGATCCGCCAGAGCACTAAG
 GACTCCCCGCTCTGGGAGTTCTTGAGCCTGGGCTACGCGCTCATGCTCTGCCCTTTCGTC
 GTGGTCCTGGGCGGCATGTTCTTCTCGCCACTGCGCTCTTCTTCGTCAGCGACCGCGCC
 AGGGCTGAGCAGCACCTGGGGGAGAGACGGGCGGGGGTCAAGGTGGTGCATCAGCGGGGG
 CCGGGCCCGGGCACTGCTCTGGCACATCGTGTCTGTGGGGGCCAGCTGACCGGAGGTGCTG
 GGCAGGGACCTCGTCAGCCCCAGGGGGAGATGGGAGAGCCAGGGGTGGGGAGAGGAGAG
 AGAGAGGAGTAAAGAGGAAAGGAGAAAGAAGTCAGAAAGTAAGAGGAAGGGGAGGGGCC
 CAGCTTTGAAAACCACTAAGTCCAGAGACAAACCCAAGTCTGGATCCACCAGACACCCCC
 GTGGCTCCACAGCTCCAGGCTGACCCTGGCACTGGGCCTCAGGGCTGGACCCAGCAA
 CCAGTGGGTGCACTGAGTGCATGGGAGGTCTGTACCCTCCCCGCCCCACCCAGGGCAGG
 GCTCACGGTGGCTATCACGGTCCCTGCTTCC

In a search of public sequence databases, the NOV19 nucleic acid sequence, located on chromosome 17 has 290 of 431 bases (67%) identical to a gb:GENBANK-

ID:E12646|acc:E12646.1 mRNA from Homo sapiens (cDNA encoding cell growth inhibiting factor). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV19 polypeptide (SEQ ID NO:48) encoded by SEQ ID NO:47 has 566 amino acid residues and is presented in Table 19B using the one-letter amino acid code.

Signal P, Psort and/or Hydropathy results predict that NOV19 has no signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.6000.

Table 19B. Encoded NOV19 protein sequence (SEQ ID NO:48).

MMCLECASAAGGAEEEEADAERRRRRGAQRGAGGSGCCGARGAGGAGVSAAGDEVQTL
 SGSVRRAPTGPPTGTPGCAATAKPGQAQPKPASLGRGRGAAAAILSLGNVNLNYLDY
 TVAGVLLDIQQHFGVKDRGAGLLQSVFICSFMAAPIFGYLGDRFNKRVILSCGIFFWSA
 VTFSSSFIPQQYFWLLVLSRGLVGIGEASYSTIAPTIIIGDLFTKNTRLMLSVFYFAIPL
 GSGGLGYITGSSVKQAAGDWHWALRVSPVLGMITGTLLILVPATKRHADQLGDQLKART
 SWLRDMKALIRNRSYVFSSLATSAVSFATGALGMWIPLYLHRAQVQKTAETCNSPPCGA
 KDSLIFGAITCFTGFLGVVTGAGATRWCRCLKTQRADPLVCAVGMGLSAIFICLIFVAKS
 SIVGAYICIFVGETLLFSNWAITADILMYVVIPTRRATAVALQSFTSHLLGDAGSPYLIG
 FISDLIRQSTKDSPLWEFLSLGYALMLCPFVVVLGGMFFLATALFFVSDRARAQHLGER
 RAGVRVVHQRGPGPGTALAHRVVGAS

A search of sequence databases reveals that the NOV19 amino acid sequence has 268 of 495 amino acid residues (54%) identical to, and 330 of 495 amino acid residues (66%) similar to, the 528 amino acid residue ptnr:TREMBLNEW-ACC:AAG43830 protein from Homo sapiens (Human) (SPINSTER-LIKE PROTEIN). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV is expressed in at least brain and heart. This information was derived by determining the tissue sources of the sequences that were included in the invention including

but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV19 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 19C.

5

Table 19C. BLAST results for NOV19					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
<u>gi 15079262 gb AAH11467.1 AAH11467</u> (BC011467)	protein for IMAGE:3154539) [Mus musculus]	590	514/539 (95%)	523/539 (96%)	0.0
<u>gi 18448989 gb AAL69987.1 AF465772.1</u> (AF465772)	not really started [Danio rerio]	506	248/431 (57%)	312/431 (71%),	e-133
<u>gi 13544043 gb AAH06156.1 AAH06156</u> (BC006156)	(protein for IMAGE:3627317) [Homo sapiens]	524	266/478 (55%)	327/478 (67%),	e-129
<u>gi 14042968 ref NP114427.1 </u> (NM_032038)	spinster-like protein [Homo sapiens]	528	268/495 (54%)	330/495 (66%),	e-129
<u>gi 12963795 ref NP076201.1 </u> (NM_023712)	spinster-like protein [Mus musculus]	528	266/510 (52%)	330/510 (64%),	e-126

Table 19D-E lists the domain descriptions from DOMAIN analysis results against NOV19. This indicates that the NOV sequence has properties similar to those of other proteins known to contain this domain.

10

Table 19D. Domain Analysis of NOV19
<u>gnl Pfam pfam00083</u> , sugar_tr, Sugar (and other) transporter.
CD-Length = 447 residues, 28.2% aligned
Score = 55.1 bits (131), Expect = 1e-08

Table 19E. Domain Analysis of NOV19

gnl|Pfam|pfam03137, OATP_C, Organic Anion Transporter Polypeptide (OATP) family, C-terminus. This family consists of several eukaryotic Organic-Anion-Transporting Polypeptides (OATPs). Several have been identified mostly in human and rat. Different OATPs vary in tissue distribution and substrate specificity. Since the numbering of different OATPs in particular species was based originally on the order of discovery, similarly numbered OATPs in humans and rats did not necessarily correspond in function, tissue distribution and substrate specificity (in spite of the name, some OATPs also transport organic cations and neutral molecules). Thus, Tamai et al. initiated the current scheme of using digits for rat OATPs and letters for human ones. Prostaglandin transporter (PGT) proteins (e.g. Pfam:Q92959) are also considered to be OATP family members. In addition, the methotrexate transporter OATK (Pfam:P70502) is closely related to OATPs. This family aligns residues towards the C-terminus. The family OATP_N aligns residues from similar proteins towards the N-terminus. This family also includes several predicted proteins from *Caenorhabditis elegans* and *Drosophila melanogaster*. This similarity was not previously noted. Note: Members of this family are described (in the Swiss-Prot database) as belonging to the SLC21 family of transporters.

CD-Length = 375 residues, 18.9% aligned

Score = 37.7 bits (86), Expect = 0.002

- NOV19 is a homolog of the spinster-like proteins in human and mouse. Spinster is a novel membrane protein in *Drosophila*, mutants of which exhibit accumulation of ceroid lipofuscin and neural degeneration (See Nakano et al., Genbank entry for AAG43830.1).
- 5 Accumulation of ceroid lipofuscin occurs in several hereditary disorders that are probably related to lysosomal storage defects. The pigment makes fibroblasts in vitro more susceptible to oxidative stress, leading to apoptosis (See Terman et al., Exp Gerontol 1999 Sep;34(6):755-70). It is also a component of atherogenic lesions in arteries (See Hoffe and Hoppe, Curr Opin Lipidol 1995 Oct;6(5):317-25). In neurons, accumulation of the pigment leads to
- 10 neurodegeneration, even leading to death in some cases (See Dyken and Wisniewski, Am J Med Genet 1995 Jun 5;57(2):150-4). This is seen both in inherited forms of human ceroid lipofuscinoses (See Jagadha et al., Acta Neuropathol (Berl) 1988;75(3):233-40) and in a cathepsin-D-deficient mouse model (See Koike et al., J Neurosci 2000 Sep 15;20(18):6898-906). In at least one case, neuronal deposition of ceroid lipofuscin was also correlated with
- 15 accumulation in the myocardium (See Jay and Haslam, J Inherit Metab Dis 1995;18(3):359-60).

The disclosed NOV19 nucleic acid of the invention encoding a spinster-like protein includes the nucleic acid whose sequence is provided in Table 19A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed

from the corresponding base shown in Table 19A while still encoding a protein that maintains its spinster-like activities and physiological functions, or a fragment of such a nucleic acid.

The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids

5 just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be
10 used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 33 percent of the bases may be so changed.

The disclosed NOV19 protein of the invention includes the spinster-like protein whose sequence is provided in Table 19B. The invention also includes a mutant or variant protein
15 any of whose residues may be changed from the corresponding residue shown in Table 19B while still encoding a protein that maintains its spinster-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 46 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or
20 (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this spinster-like protein (NOV19) may function as a member of a “spinster family”. Therefore, the NOV19 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The
25 potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

30 The NOV19 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the spinster-like protein (NOV19) may be useful in gene therapy, and the spinster-like protein (NOV19) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the

compositions of the present invention will have efficacy for treatment of patients suffering from cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, transplantation, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neurodegeneration, cancer, tissue degeneration, bacterial/viral/parasitic infection, or other pathologies or conditions. The NOV19 nucleic acid encoding the spinster-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV19 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV19 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV19 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV20

A disclosed NOV20 nucleic acid of 752 nucleotides (also referred to as CG57695-01) encoding a casein-like protein is shown in Table 20A. The start and stop codons are in bold letters.

Table 20A. NOV20 nucleotide sequence (SEQ ID NO:49).

CAAACAAC**ATGA**AGTTCCTTCATCTTTACCTGCCTTTTGGCTGTTGCTCTGGCACATCAT
GAACTTAAACACGTTTACAAGAAAAAACAACAACGTGAGTGACAAATACAGAAAT
GTGAAAAACCAGATTTCTTCCTCAGGAGGACAAAGTTAGAGGTAATTTTCATTCAAAT
AAAATAAAATTAATACCACTTAGTAGTGTTTTATTTTGTATATTTGTATACAAATTAAT
TTTTTTTCTTACCAGGAAGTTAAGCACACTGTGGATCAAAAGCACTACGTAAAACAACCTG
AACAAATCAACCCATTTTATCAGAAGTGGAACCTCCTCCCATTTCTTCAGGTTCTTTAT
CAACATCAGATTTTATAAACCAGGAGATCAGCATAAGACAAGTGCTACCCCTTTGTT
CCCACTAAATATATACAGTGGCCAGGCTCAGTGGCTCAGGCCTTCTTGTTTTATTCTTT
AAGGAAACACCAAAAAAGACTGTTGATATGGTAAAGTATTGTTCTATCAGAAAACTGAG
CTGACTGAAGAAGAAAAGAATGACCAAAAACATCTGAACAAAATCAACAGTATTATCAG
TTCACCTTGCCCCAATATGTAAAAGCTGTTTATCAATATCACAAAATTATGAAACCATGG

AAAAACATGAAGACAAATGCTTACCAAGTTATCCCCACTCTGGTGAGTGCTCTCTTTT
TTTGCAACTTAAAAATAGTTATCTGTTGTGCT

In a search of public sequence databases, the NOV20 nucleic acid sequence, located on chromosome 4 has 291 of 445 bases (65%) identical to a gb:GENBANK-
ID:CHI249995|acc:AJ249995.1 mRNA from Capra hircus (Capra hircus mRNA for alpha s2-
casein (cns1s2 gene)). Public nucleotide databases include all GenBank databases and the
GeneSeq patent database.

The disclosed NOV20 polypeptide (SEQ ID NO:50) encoded by SEQ ID NO:49 has 240 amino acid residues and is presented in Table 20B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV20 has a signal peptide and is
likely to be localized extracellularly with a certainty of 0.5140.

Table 20B. Encoded NOV20 protein sequence (SEQ ID NO:50).

MKFFIFTCLLAVALAHHELKHVYKKKTNNNVSDKYRNVKNQISSPQEDKVRGNFHSNKIK
LIPLSSVLFLYICIQINFFSYQEVKHTVDQKHVYKQLNKINPFYQKWNFLPFLQVPYQH
IFINPGDQHKTSVYPFVPTKYIQWPGSVAQAFYFYSFKETPKKTVDMMVKYCFYQKTELTE
EEKNDQKHLNKNQYYQFTLPQYVKAQYQYHKIMKPWKNMKTNAYQVIPTLVSAFLFAT

A search of sequence databases reveals that the NOV20 amino acid sequence has 112 of 232 amino acid residues (48%) identical to, and 142 of 232 amino acid residues (61%) similar to, the 235 amino acid residue ptnr:pir-id:A48383 protein from pig (alpha s2-casein).
Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV20 is expressed in at least lung, testis, and b-cell. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV20 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 20C.

Table 20C. BLAST results for NOV20

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Pos itives (%)	Expect
gi 477186 pir A48383	alpha s2-casein - pig	235	111/234 (47%)	141/234 (59%)	2e-37

gi 729044 sp P39036 CAS2_PIG	Alpha-S2 casein precursor	235	109/234 (46%)	138/234 (58%)	2e-35
gi 115658 sp P04654 CAS2_SHEEP	Alpha-S2 casein precursor	223	100/227 (44%)	128/227 (56%)	3e-28
gi 416751 sp P33049 CAS2_CAPHI	Alpha-S2 casein precursor (Alpha-S2-CN)	223	96/227 (42%)	125/227 (54%)	4e-28
gi 423305 pir S33881	alphas2-casein B - goat	223	96/227 (42%)	125/227 (54%) ,	4e-28

Table 20D lists the domain descriptions from DOMAIN analysis results against NOV20. This indicates that the NOV20 sequence has properties similar to those of other proteins known to contain this domain.

Table 20D. Domain Analysis of NOV20	
gnl Pfam pfam00363, caseins, Casein.	
CD-Length = 84 residues, 71.4% aligned	
Score = 44.3 bits (103), Expect = 8e-06	

NOV20 has homology to pig alpha S casein. Caseins are the major protein constituent of milk. Caseins can be classified into two families; the first consists of the kappa-caseins, and the second groups the alpha-s1, alpha-s2, and beta-caseins. The alpha/beta caseins are a rapidly diverging family of proteins. However two regions are conserved: a cluster of phosphorylated serine residues and the signal sequence. Alpha-s2 casein is known as epsilon-casein in mouse, gamma-casein in rat and casein-A in guinea pig. Alpha-s1 casein is known as alpha-casein in rat and rabbit and as casein-B in guinea-pig.

Milk casein can be separated by urea starch electrophoresis into 3 regions, alpha, beta (115460), and kappa (601695) casein. Alpha and beta variants are present in the human population. Voglino and Ponzzone (See Voglino, G. F.; Ponzzone, A.: Nature N.B. 238: 149, 1972) postulated 2 biallelic systems. In Italy the frequency of the 2 alpha alleles was 0.908 and 0.092; 2 beta alleles had a frequency of 0.678 and 0.322.

Fujiwara et al. (See Fujiwara, Y.et al., Hum. Genet. 99: 368-373, 1997) found that the human alpha-S1, beta-, and kappa-casein genes are closely linked and arranged in that order. By fluorescence in situ hybridization, they demonstrated that the casein gene family is localized to 4q21.1. Rijnkels et al. (See Rijnkels, M., et al., Mammalian Genome 8: 285-286, 1997) concluded that the human 'locus' comprises at least 4 casein genes: 3 genes encoding calcium-sensitive, casein-like genes, and 1 kappa-casein gene, in the order alpha-s1--beta--

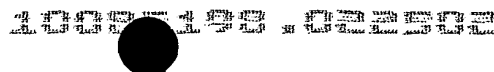
alpha-s2--kappa. The approximate size of the human casein gene locus is 350 kb. Chen et al. (See Chen, C.-S. et al., Cytogenet. Cell Genet. 69: 260-265, 1995.) suggested that the casein cluster is located within 700 kb of the albumin (103600) gene cluster, which is located on 4q13.

5 The disclosed NOV20 nucleic acid of the invention encoding a casein-like protein includes the nucleic acid whose sequence is provided in Table 20A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 20A while still encoding a protein that maintains its casein-like activities and physiological functions, or a fragment of such a nucleic acid. The
10 invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar
15 phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 35 percent of the bases may be so changed.

20 The disclosed NOV20 protein of the invention includes the casein-like protein whose sequence is provided in Table 20B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 20B while still encoding a protein that maintains its casein-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 52
25 percent of the residues may be so changed.

 The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

 The above defined information for this invention suggests that this casein-like protein (NOV20) may function as a member of a "casein family". Therefore, the NOV20 nucleic acids
30 and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation),



research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV20 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the casein-like protein (NOV20) may be useful in gene therapy, and the casein-like protein (NOV20) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from fertility, hypogonadism, systemic lupus erythematosus, autoimmune disease, asthma, emphysema, scleroderma, allergy, ARDS, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, graft versus host disease (GVHD), lymphoedema, or other pathologies or conditions. The NOV20 nucleic acid encoding the casein-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV20 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV20 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV20 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV21

A disclosed NOV21 nucleic acid of 1704 nucleotides (also referred to as CG57654-01) encoding a gamma-aminobutyric acid receptor-like protein is shown in Table 21A. The start and stop codons are in bold letters.

Table 21A. NOV21 nucleotide sequence (SEQ ID NO:51).

<p>CCTGACGCTTTGATGGTATCTGCAAGCGTTTTTGCTGATCTTATCTCTGCCCCCTGAATA TTAATTCCCTAATCTGGTAGCAATCCATCTCCCCAGTGAAGGACCTACTAGAGGCAGGTG GGGGGAGCCACCATCAGATCATCAAGCATAAGAATAATACAAAGGGGAGGGATTCTTCTG CAACCAAGAGGCAAGAGGCGAGAGAAGGAAAAAAAAAAAAAAAAAGCGATGAGTTCACCAA TATATGGAGCACAGGAAGCTCAGTCTACTCGACTCCTGTATTTTCACAGAAAATGACGGT</p>
--

GTGGATTCTGCTCCTGCTGTCGCTCTACCCTGGCTTCACTAGCCAGAAATCTGATGATGA
CTATGAAGATTATGCTTCTAACAAAACATGGGTCTTGACTCCAAAAGTTCCTGAGGGTGA
TGTCACCTGTCATCTTAAACAACCTGCTGGAAGGATATGACAATAAACTTCGGCCTGATAT
AGGAGTGAAGCCAACGTTAATTCACACAGACATGTATGTGAATAGCATTGGTCCAGTGAA
CGCTATCAATATGGAATACACTATTGATATATTTTTTGCACAAATGTGGTATGACAGACG
TTTGAAATTTAACAGCACCATTAAAGTCCTCCGATTGAACAGCAACATGGTGGGGAAAAT
CTGGATTCAGACACTTTCTTCAGAAATTCCAAAAAGCTGATGCACACTGGATCACCAC
CCCCAACAGGATGCTGAGAATTTGGAATGATGGTCGAGTGTCTTACTCCCTAAGGTTGAC
AATTGATGCTGAGTGCCAATTACAATTGCACAATTTTCCAATGGATGAACACTCCTGCCC
CTTGAGTTCCTCCAGTTATGGCTATCCACGTGAAGAAATTGTTTATCAATGGAAGCGAAG
TTCTGTTGAAGTGGGCGACACAAGATCCTGGAGGCTTTATCAATTCTCATTGTGTGGTCT
AAGAAATACCACCGAAGTAGTGAAGACAACCTCCGGAGATTATGTGGTCATGTCTGTCTA
CTTTGATCTGAGCAGAAGAATGGGATACTTTACCATCCAGACCTATATCCCCTGCACACT
CATTGTCTGCTATCCTGGGTGTCTTTCTGGATCAATAAGGATGCTGTTCCAGCCAGAAC
ATCTTTAGGTATCACCCTGTCTGACAATGACCACCCTCAGCACCATTGCCCGGAAATC
GCTCCCCAAGGTCTCCTATGTACAGCGATGGATCTCTTTGTATCTGTTTGTTCATCTT
TGTCTTCTGCTCTGGTGGAGTATGGCACCTTGCAATTATTTTGTGAGCAACCGGAAACC
AAGCAAGGACAAAGATAAAAAGAAGAAAAACCTCTTCTTCGGATGTTTTCTTCAAGGC
CCCTACCATTGATATCCGCCCAAGATCAGCAACCATTCAAATGAATAATGTACACACCT
TCAAGAGAGAGATGAAGAGTACGGCTATGAGTGTCTGGACGGCAAGGACTGTGCCAGTTT
TTTCTGCTGTTTGAAGATTGTGCAACAGGAGCTTGGAGACATGGGAGGATACATATCCG
CATTGCCAAAATGGACTCCTATGCTCGGATCTTCTTCCCCACTGCCTTCTGCTGTTTAA
TCTGGTCTATTGGGTCTCCTACCTCTACCTGTGAGGAGGTATGGGTTTTACTGATATGGT
TCTTATTCACCTGAGTCTCATGGAG

In a search of public sequence databases, the NOV21 nucleic acid sequence, located on chromosome 5 has 1379 of 1398 bases (98%) identical to a gb:GENBANK-ID:HSGABAAS|acc:X15376.1 mRNA from Homo sapiens (Human mRNA for GABA-A receptor, gamma 2 subunit). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV21 polypeptide (SEQ ID NO:52) encoded by SEQ ID NO:51 has 475 amino acid residues and is presented in Table 21B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV21 has a signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.6000. The most likely cleavage site for a NOV21 peptide is between amino acids 39 and 40.

Table 21B. Encoded NOV21 protein sequence (SEQ ID NO:52).

MSSPNIWSTGSSVYSTPVFSQKMTVWILLLLSLYPGFTSQKSDDDYEDYASNKTWVLTPK
VPEGDVTIVILNNLLEGYDNKLRPDIGVKPTLIHTDMYVNSIGPVNAINMEYTIIDIFFAQM
WYDRRLKFNSTIKVLRNLNSNMVGKIWIPDTFFRNSKKADAHWITTPNRMLRIWNDGRVLY
SLRLTIDAECQLQLHNFPMDEHSCPLEFSSYGYPREEIVYQWKRSSVEVGDRSWRLYQF
SFVGLRNTTEVVKTTSGDYVMSVYFDLSRRMGYFTIQTYIPCTLIIVVLSWVSFWINKDA
VPARTSLGITTIVLTMTTLSTIARKSLPKVS YVTAMD L FVSVC FIFVFSALVEYGTLHYFV
SNRKPSKDKDKKKKNPLLRMF S F KAPTIDIRPSATI QMNATHLQERDEEYGYECLDGK
DCASFFCCCFEDCRTGAWRHGRIHIRIAKMDSYARIFFP TAFCLFNLVYVWSYLYL

A search of sequence databases reveals that the NOV21 amino acid sequence has 467 of 475 amino acid residues (98%) identical to, and 467 of 475 amino acid residues (98%)

similar to, the 467 amino acid residue ptnr:SWISSNEW-ACC:P18507 protein from Homo sapiens (Human) (GAMMA-AMINOBUTYRIC-ACID RECEPTOR GAMMA-2 SUBUNIT PRECURSOR (GABA(A) RECEPTOR)). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

5 NOV21 is expressed in at least Adrenal Gland/Suprarenal gland, Brain, Hippocampus, Pituitary Gland, and Right Cerebellum. Expression information was derived from the tissue sources of the sequences that were included in the derivation of the sequence of CG57654-01. The sequence is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:HSGABAAS|acc:X15376.1) a closely related
10 Human mRNA for GABA-A receptor, gamma 2 subunit homolog in species Homo sapiens: fetal brain. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

15 The disclosed NOV21 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 21C.

Table 21C. BLAST results for NOV21					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
<u>gi 4557611 ref NP_00807.1 </u> (NM_000816)	gamma-aminobutyric acid A receptor, gamma 2 precursor [Homo sapiens]	467	467/475 (98%)	467/475 (98%)	0.0
<u>gi 5738138 gb AAD50273.1 </u> (AF165124)	gamma-aminobutyric acid A receptor gamma 2 [Homo sapiens]	467	465/475 (97%)	466/475 (97%)	0.0
<u>gi 120784 sp P22300 GAC2 BOVIN</u>	GAMMA-AMINOBUTYRIC-ACID RECEPTOR GAMMA-2 SUBUNIT PRECURSOR (GABA (A) RECEPTOR)	475	467/475 (98%)	470/475 (98%)	0.0
<u>gi 108682 pir B39272</u>	gamma-aminobutyric acid receptor A gamma-2L chain - bovine	475	467/475 (98%)	470/475 (98%)	0.0
<u>gi 6679915 ref NP_032099.1 </u> (NM_008073)	gamma-aminobutyric acid (GABA-A) receptor, subunit gamma 2 [Mus musculus]	474	469/475 (98%)	471/475 (98%)	0.0

Table 21D lists the domain descriptions from DOMAIN analysis results against NOV21. This indicates that the NOV21 sequence has properties similar to those of other proteins known to contain this domain.

Table 21D. Domain Analysis of NOV21	
gnl Pfam pfam02931, Neur_chan_LBD, Neurotransmitter-gated ion-channel ligand binding domain. This family is the extracellular ligand binding domain of these ion channels. This domain forms a pentameric arrangement in the known structure.	
CD-Length = 200 residues, 91.5% aligned	
Score = 170 bits (431), Expect = 1e-43	

5

Neurotransmission effected by GABA (gamma-aminobutyric acid) is predominantly mediated by a gated chloride channel intrinsic to the GABAA receptor. This heterooligomeric receptor exists in most inhibitory synapses in the vertebrate central nervous system (CNS) and can be regulated by clinically important compounds such as benzodiazepines and barbiturates.

10 The primary structures of GABAA receptor alpha- and beta-subunits have been deduced from cloned complementary DNAs. Co-expression of these subunits in heterologous systems generates receptors which display much of the pharmacology of their neural counterparts, including potentiation by barbiturates. Conspicuously, however, they lack binding sites for, and consistent electrophysiological responses to, benzodiazepines. (See Pritchett et al. *Nature*

15 1989;338:582-5) reported the isolation of a cloned cDNA encoding a new GABAA receptor subunit, termed gamma 2, which shares approximately 40% sequence identity with alpha- and beta-subunits and whose messenger RNA is prominently localized in neuronal subpopulations throughout the CNS. Importantly, coexpression of the gamma 2 subunit with alpha 1 and beta 1 subunits produces GABAA receptors displaying high-affinity binding for central

20 benzodiazepine receptor ligands.

The disclosed NOV21 nucleic acid of the invention encoding a gamma-aminobutyric acid receptor-like protein includes the nucleic acid whose sequence is provided in Table 21A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 21A while still

25 encoding a protein that maintains its gamma-aminobutyric acid receptor-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements

thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 2 percent of the bases may be so changed.

The disclosed NOV21 protein of the invention includes the gamma-aminobutyric acid receptor-like protein whose sequence is provided in Table 21B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 21B while still encoding a protein that maintains its gamma-aminobutyric acid receptor-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 2 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this gamma-aminobutyric acid receptor-like protein (NOV21) may function as a member of a "gamma-aminobutyric acid receptor family". Therefore, the NOV21 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV21 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the gamma-aminobutyric acid receptor-like protein (NOV21) may be useful in gene therapy, and the gamma-aminobutyric acid receptor-like protein (NOV21) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from adrenoleukodystrophy, congenital adrenal hyperplasia, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease,

stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neurodegeneration, endocrine dysfunctions, diabetes, obesity, growth and reproductive disorders, or other pathologies or conditions. The NOV21 nucleic acid encoding the gamma-aminobutyric acid receptor-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV21 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV21 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV21 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV22

A disclosed NOV22 nucleic acid of 1602 nucleotides (also referred to as 57724-01) encoding a carboxylesterase-like protein is shown in Table 22A. The start and stop codons are in bold letters.

Table 22A. NOV22 nucleotide sequence (SEQ ID NO:53).

ATGCGGCTGCACAGACTTCACGCGCGGCCGAGCGCGGTGGCCTCTGCTGCTTCTGATGCTGT
GTGGGCCCGAAGTTGCTCAGCCTGAAGTAGACACCACCTGGGTCTGTGCGAGGCCGAGGTGGGCGT
GAAGGGCACAGACCGCCTTGTGAATGTCTTTCTGGGCATTCCATTTGCCAGCCGCCACTGGGCCCTGAC
CGTTTCTCAGCCCCACACCCAGCACAGCCCTGGGAGGGTGTGCGGGATGCCAGCACTGCGCCCCAATGT
GCCTACAAGACGTGATGAACAGCAGCAGATTTGTCTCAACGGAAAACAGCAGATCTTCTCCGTTTCAGA
GGACTGCCTGGTCTCAACGTCTATAGCCAGCTGAGGTCTATGGTATGGGTCCATGGAGGCGCTCTGATA
ACTGGCGCTGCCACCTCCTACGATGGATCAGCTCTGGCTGCCTATGGGGATGTGGTCTGTTACAGTCC
AGTACCGCCTTGGGGTCTTGGCTTCTTCAGCACTGGAGATGAGCATGCACCTGGCAACCAGGGCTTCT
AGATGTGGTAGCTGCTTTGCGCTGGGTGCAAGAAAACATCGCCCCCTTCGGGGGTGACCTCAACTGTGTC
ACTGTCTTTGGTGGATCTGCCGGTGGGAGCATCATCTCTGGCCTGGTCTGTCCCCAGTGGCTGCAGGGC
TGTTCCACAGAGCCATCACACAGAGTGGGGTCAACACCCAGGGATCATCGACTCTCACCTTGGCC
CCTAGCTCAGAAAATCGCAAACACCTTGGCCTGCGAGCTCCAGCTCCCCGGCTGAGATGGTGCAGTGCCTT
CAGCAGAAAAGAAGGAGAAGAGCTGGTCTTAGCAAGAAGCTGAAAAATACTATCTATCTCTCACCGTTG
ATGGCACTGTCTTCCCCAAAAGCCCCAAGGAACCTCTGAAGGAGAAGCCCTTCCACTCTGTGCCCTTCT
CATGGGTGTCAACAACCATGAGTTCAGCTGGCTCATCCCCGGGACCAAGGTGATGCGTGTGTCCAACAAG
ATGATCATGAAGTTCGCTAAACCGGCAGGCGATGAGAAAGGAAACCATCACTAAGATGCTCTGGAGTA
CCCGCACCTGTGGAGCATGACTGGAAGATGCTACGAAACCGTATGATGGACATAGTTCAAGATGCCAC
TTTCGTGTATGCCCACTGCAGACTGCTCACTACACCGAGATGCCGGCTCCCTGTCTACCTGTATGAA
TTTGAGCACCACGCTCGTGGAATAATCGTCAAACCCCGCACTGATGGGGCAGACCATGGGGATGAGATGT
ACTTCTCTTTGGGGGCCCCCTTCGCCACAGGCCTTTCATGGGTAAGGAGAAGGCACCTAGCCTCCAGAT
GATGAAATACTGGGCCAACTTTGCCCGCACAGGAAACCCCAATGATGGGAATCTGCCCTGCTGGCCACGC
TACAACAAGGATGAAAAGTACCTGCAGCTGGATTTTACCACAAGAGTGGGCATGAAGCTCAAGGAGAAGA

AGATGGCTTTTGGATGAGTCTGTACCAGTCTCAAAGACCTGAGAAGCAGAGGCAATTCTAA

In a search of public sequence databases, the NOV22 nucleic acid sequence, located on chromosome 16 has 695 of 735 bases (94%) identical to a gb:GENBANK-ID:AK000105|acc:AK000105.1 mRNA from Homo sapiens (Homo sapiens cDNA FLJ20098 fis, clone COL04537, highly similar to ESTM_MOUSE LIVER CARBOXYLESTERASE PRECURSOR). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV22 polypeptide (SEQ ID NO:54) encoded by SEQ ID NO:53 has 533 amino acid residues and is presented in Table 22B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV22 has a signal peptide and is likely to be localized extracellularly with a certainty of 0.7953. The most likely cleavage site for a NOV22 peptide is at amino acid position 29.

Table 22B. Encoded NOV22 protein sequence (SEQ ID NO:54).

MRLHRLHARPSAVACGLLLLLMLCGPEVAQPEVDTTLGRVRGRQVGKGTDRLVNVFLGI
PFAQPPLGPDRFSAPHPAQPWEGVRDASTAPPMCLQDVMNSSRFVLNGKQQIFSUSEDCL
VLNVYSPAEMVWVHGGALITGAATSYDGSALAAYGDVVVTVQYRLGVLGFFSTGDEHA
PGNQGFLDVVAALRWQENIAPFGDLNCVTVFSGSAGGSIIISGLVLSPLAAGLFHRAIT
QSGVITTPGIIDSHPWPLAQKIANTLACSSSSPAEMVQCLQQKEGEELVLSKKLKNTIYP
LTVDGTVPFKSPKELLKEKPFHSVPFLMGVNNHEFSWLI PGTKVMRVS NKMIMKFP LN RQ
AMRKETITKMLWSTRTLLEHDWKMLRNRMDIVQDATFVYATLQTAHYHRDAGLPVYLYE
FEHHARGIIVKPRTDGADHGDEMYFLFGGPFATGLSMGKEKALSLQMMKYWANFARTGNP
NDGNLPCWPRYNKDEKYLQLDFTTRVGMKLKEKKMAFWMSLYQSQRPEKQRQF

A search of sequence databases reveals that the NOV22 amino acid sequence has 296 of 544 amino acid residues (54%) identical to, and 373 of 544 amino acid residues (68%) similar to, the 554 amino acid residue ptnr:SWISSPROT-ACC:Q63880 protein from Mus musculus (Mouse) (LIVER CARBOXYLESTERASE PRECURSOR (EC 3.1.1.1) (ES-MALE) (ESTERASE-31)). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV22 is expressed in at least liver, colon, small intestine, kidney, pancreas, brain, and plasma. Expression information was derived from the tissue sources of the sequences that were included in the derivation of the sequence of CG57724-01. The sequence is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:AK000105|acc:AK000105.1) a closely related Homo sapiens cDNA FLJ20098 fis, clone COL04537, highly similar to ESTM_MOUSE LIVER CARBOXYLESTERASE PRECURSOR homolog in species Homo sapiens : liver, colon,

small intestine, kidney, pancreas, brain, and plasma. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

- 5 The disclosed NOV22 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 22C.

Table 22C. BLAST results for NOV22					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
<u>gi 17512361 gb AAH19147.1 AAH19147</u> (BC019147)	protein for MGC:29382) [Mus musculus]	568	293/547 (53%)	370/547 (67%)	e-156
<u>gi 2494382 sp Q63880 ESTM MOUSE</u>	LIVER CARBOXYLESTERASE PRECURSOR (ES- MALE) (ESTERASE- 31)	554	297/560 (53%)	369/560 (65%)	e-155
<u>gi 14789873 gb AAH10812.1 AAH10812</u> (BC010812)	(protein for IMAGE:4211034) [Mus musculus]	524	272/538 (50%)	346/538 (63%)	e-140
<u>gi 730714 sp Q04791 SASB ANAPL</u>	FATTY ACYL-COA HYDROLASE PRECURSOR, MEDIUM CHAIN (THIOESTERASE B)	557	= 252/547 (46%)	346/547 (63%)	e-136
<u>gi 57554 emb CAA46391.1 X65296</u>	carboxylesterase [Rattus rattus]	565	(42%)	336/559 (59%)	e-124

- 10 Table 22D lists the domain descriptions from DOMAIN analysis results against NOV22. This indicates that the NOV22 sequence has properties similar to those of other proteins known to contain this domain.

Table 22D. Domain Analysis of NOV22
<u>gnl Pfam pfam00135</u> , COesterase, Carboxylesterase
CD-Length = 532 residues, 93.0% aligned
Score = 374 bits (960), Expect = 8e-105

- 15 The mammalian carboxylesterases (EC 3.1.1.1) comprise a multigene family, the gene products of which are localized in the endoplasmic reticulum (ER) and cytosol of many tissues. These enzymes efficiently catalyze the hydrolysis of a variety of ester- and amide-containing chemicals, as well as drugs (including prodrugs) to the respective free acids. They are involved in detoxification or metabolic activation of various drugs, environmental toxicants, and carcinogens. Carboxylesterases also catalyze the hydrolysis of endogenous

compounds such as short- and long-chain acylglycerols, long-chain acylcarnitine, and long-chain acyl-CoA esters. Multiple isozymes of hepatic microsomal carboxylesterases exist in various animal species, and some of these isozymes are involved in the metabolic activation of certain carcinogens and are associated with hepatocarcinogenesis.

5 Several studies have shown that various carboxylesterases are present in a wide variety of organs and tissues of many mammalian species; the highest hydrolase activity occurs in the liver. Humans express carboxylesterase in the liver, small intestine, brain, stomach, colon, pancreas, kidney, macrophages, monocytes, and plasma. Carboxylesterases, in addition to the metabolism of exogenous compounds, have been shown to hydrolyze endogenous fatty acid
10 esters of steroids in both rat pancreas and kidney. The nonspecific esterases found in brain appear to be present only in the central nervous system, and four unique carboxylesterases have been isolated from human brain extract. Carboxylesterase activity of is found predominantly in the microsomal fraction, although significant carboxylesterase activity is present in the lysosomal fraction, and the lysosomes contribute substantially to the general
15 esterolytic capacity of liver. The microsomal and lysosomal enzymes can be differentiated on the basis of both substrate specificity and structure and are considered to belong to separate classes. Carboxylesterase activity is also found in the cytosolic fraction of brain and in the plasma. Carboxylesterase is present in the plasma, but it is most likely synthesized in liver and then secreted into the circulation via the Golgi apparatus.

20 Human liver and plasma carboxylesterase activates lovastatin, and there are a significant number of additional drugs and endogenous compounds that are substrates of carboxylesterases, e.g. dipivefrin hydrochloride, carbonates, cocaine, salicylates, capsaicin, palmitoyl-coenzyme A, haloperidol, imidapril, pyrrolizidine alkaloids, and steroids.

 The disclosed NOV22 nucleic acid of the invention encoding a carboxylesterase-like
25 protein includes the nucleic acid whose sequence is provided in Table 22A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 22A while still encoding a protein that maintains its carboxylesterase-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are
30 complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or

5 The disclosed NOV22 protein of the invention includes the carboxylesterase-like protein whose sequence is provided in Table 22B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 22B while still encoding a protein that maintains its carboxylesterase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, 10 up to about 46 percent of the residues may be so changed.

The above defined information for this invention suggests that this carboxylesterase-like protein (NOV22) may function as a member of a “carboxylesterase family”. Therefore, the NOV22 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

148

antibodies that bind immuno-specifically to the novel NOV22 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV22 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV23

A disclosed NOV23 nucleic acid of 996 nucleotides (also referred to as CG57730-01) encoding a MAT-1-like protein is shown in Table 23A. The start and stop codons are in bold letters.

Table 23A. NOV23 nucleotide sequence (SEQ ID NO:55).

CTCTCGAATTCCCCACCCACCTGTACTCTGGAGAGACTGTGCTGGGAACATGTACCACT
GAGCCTGAGATGGGGATGAGGGCAGAGAGAGGGGAGCCCCCTTCCACTCAGTTGTTCC
TACTCAGACTGTTGCACTCTAAACCTAGGGAGGTTGAAGAATGAGACCCTTAGGTTTTAA
CACGAATCCTGACACCACCATCTATAGGGTCCCAACTTGGTTATTGTAGGCAACCTTCCC
TCTCTCCTTGGTGAAGAACATCCCAAGCCAGAAAGAAGTTAACTACAGTGTTTTCTTTG
CACCGATCCCCACCCCAATTCAATCCCGAAGGGACTTACTTAGGAAACCTTCTTTACT
AGATATCCTGGCCCCCTGGGCTTGTGAACACCTCCTAGCCACATCACTACAGTACAGTGA
GTGACCCACAGCCTCCTGCCTACCCCAAGATGCCCCCTCCCCACCCTGACCGTGCTAACTGT
GTGTACATATATATTCTACATATATGTATATTTAAACTGCACTGCCATGTCTGCCCTTTT
TTGTGGTGTCTAGCATTAACTTATTGTCTAGGCCAGAGCGGGGTGGGAGGGGAATGCCA
CAGTGAAGGGAGTGGCAGAATCAAATTGCTACATAGTCCAAACAAAAAAGAAGGCTTTTT
CAAAAAACATTAAATTCACATGCAGTCTCAGAGACTATTAGACAAAGTTCAAGTTAGGA
GCTTTTAGGATGTGGGAGTAAACTTTAATGGGAGGGGAGGGCTGGCTGCTGGAAGAAGG
AAGAAGCCAGACTGGTTAGACAGTACTCTTAACCTCCTAGCCAGCCTAGCGTGCCCTGCC
CCTCTGGCCACTGCTGCAGACACCTGCCTTAACACACACCTCTAGGACTCCACAGTTT
TGCCTTAAAGGACCTTCCCAAGTCTCCCTTTCCCTGTCTGGCTTCTCCCTTAAGAAGAGA
GAGATACTTGTAAGATTGGGTGGGGGAATTCGAGAG

In a search of public sequence databases, the NOV23 nucleic acid sequence, located on chromosome 1q21.1 has 983 of 987 bases (99%) identical to a gb:GENBANK-ID:PEAGENE3|acc:AF153274.1 mRNA from Homo sapiens (Homo sapiens PEA15 protein (PEA15) gene, exons 3 and 4 and complete cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV23 polypeptide (SEQ ID NO:56) encoded by SEQ ID NO:55 has 74 amino acid residues and is presented in Table B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV23 has a signal peptide and is likely to be localized in the endoplasmic reticulum with a certainty of 0.5500. The most likely cleavage site for a NOV23 peptide is between amino acids 18 and 19.

Table 23B. Encoded NOV23 protein sequence (SEQ ID NO:56).	
MYIKTALPCLPFFVSSINLLSRPERGWEGNATVKGVAESNCYIVQTKKKAFSKNIKFTC	SLRDYLDKVQVRSF

A search of sequence databases reveals that the NOV23 amino acid sequence has 62 of 75 amino acid residues (82%) identical to, and 66 of 75 amino acid residues (88%) similar to, the 75 amino acid residue ptmr:SPTREMBL-ACC:Q14801 protein from Homo sapiens (Human) (HYPOTHETICAL 8.6 KDA PROTEIN). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV23 is expressed in at least ovary, testis, brain, amygdala, pancreas, colon, and stomach.. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV23 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 23C.

Table 23C. BLAST results for NOV23					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 14714659 gb AAH10469.1 AAH10469 (BC010469)	Similar to homolog of mouse MAT-1 oncogene [Homo sapiens]	74	74/74 (100%)	74/74 (100%)	1e-34
gi 7019425 ref NP_037419.1 (NM_013287)	homolog of mouse MAT-1 oncogene; Phosphoprotein enriched in astrocytes, 15kD [Homo sapiens]	75	62/75 (82%)	66/75 (87%),	5e-27
gi 6678812 ref NP_032582.1 (NM_008556)	phosphoprotein enriched in astrocytes 15; mammary transforming gene 1 [Mus musculus]	61	27/30 (90%)	27/30 (90%)	8e-9

An efficient in vitro transformation system has been developed using N-methyl-N-nitrosourea that allows the role of hormones and growth factors in mouse mammary tumorigenesis to be studied. Utilizing this system, it was reported that mammary tumors induced in vitro with N-methyl-N-nitrosourea in the presence of mammogenic hormones (progesterone and prolactin) contain predominately an activated c-Ki-ras protooncogene with a G35 --> A35 transitional mutation in the 12th codon. Mammary tumors induced in the presence of another mitogen, lithium (Li), do not have a mutation in the c-Ki-ras protooncogene. By using an expression cloning system, a plasmid clone containing a 1.75-kb cDNA insert has been isolated from this group of tumors. Nucleic acid sequence analysis of the insert reveals that it has a short open reading frame of 61 amino acids and that it does not have sequence homology with any known gene. The gene, designated MAT1, can neoplastically transform NIH 3T3 cells and also the mammary epithelial cell line TM3. Expression of this gene occurs in normal mouse tissues including mammary gland and is overexpressed in the original mammary tumors as indicated by Northern blot analysis. In vitro transcription and translation of the clone shows a protein product of 6000 Da, which agrees with the predicted open reading frame.

The disclosed NOV23 nucleic acid of the invention encoding a MAT-1-like protein includes the nucleic acid whose sequence is provided in Table 23A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 23A while still encoding a protein that maintains its MAT-1-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 1 percent of the bases may be so changed.

The disclosed NOV23 protein of the invention includes the MAT-1-like protein whose sequence is provided in Table 23B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 23B

while still encoding a protein that maintains its MAT-1-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 12 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or 5 (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this MAT-1-like protein (NOV23) may function as a member of a “MAT-1 family”. Therefore, the NOV23 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

15 The NOV23 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the MAT-1-like protein (NOV23) may be useful in gene therapy, and the MAT-1-like protein (NOV23) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the
20 compositions of the present invention will have efficacy for treatment of patients suffering from Cataract, zonular pulverulent-1; MHC class II deficiency, complementation group C; cancer, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral
25 disorders, addiction, anxiety, pain, neurodegeneration; diabetes, pancreatitis, obesity; fertility, or other pathologies or conditions. The NOV23 nucleic acid encoding the MAT-1-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV23 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV23 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV23 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in

assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV24

- 5 A disclosed NOV24 nucleic acid of 668 nucleotides (also referred to as CG57755-01) encoding a vacuolar proton-ATPase subunit H-like protein is shown in Table 24A. The start and stop codons are in bold letters.

Table 24A. NOV24 nucleotide sequence (SEQ ID NO:57).
<p>TTCGCCCTCCCGGTCATCATCTTCACCACGTTCTGGGGCCTCGTCGGCATCGCCGGGCCCC TGGTTCGTGCCGAAGGGACCCAACCGCGGAGTGATCATCACCATGCTGGTCGCCACCGCC GTCTGCTGTTACCTCTTCTGGCTCATCGCCATCCTGGCGCAGCTGAACCCCTGTTTCGGG CCCCAGCTGAAGAATGAGACCATCTGGTACGTGCGCTTCCTGTGGGAGTGACCCGCCGCC CCCGACCCAGGTGCCAGCTCTCGGAATGACTGTGGCTCCACTGTCCCTGACAACCCCTT CGTCCGGACCCTCCCCACACAACCTATGTCTGGTCACCAGCTCCCTCCTGCTGGCACCCA GAGACCCGGACCCGCAGGGCCTGCCTGGTTCCTGGAAGTCTTCCAGTCTTCCAGCCAG CCCGGGCCCTGGGGAGCCCTGGGCACAGCAGCGGCCGAGGGGATGTCCTGCTCCAATACC CGCACTGCTCTGGAGTTTGCCCTCTTTCCCAAGGAGATGCTGCTGGGGAGCTGGTATGGG TGGGGTCTTTCCCTTTACAGACGGGGCAGATGCCAGGACTCAGCCCATCCTGAGGAGGAC ACGTGTCTCATGGAGAGGGTGCTCCGGCCAGGCGGGGGAGTCGGTGCCAGTCAGCAG GACCAGGC</p>

- 10 In a search of public sequence databases, the NOV24 nucleic acid sequence has 169 of 230 bases (73%) identical to a gb:GENBANK-ID:AF258614|acc:AF258614.1 mRNA from Canis familiaris (Canis familiaris vacuolar proton-ATPase subunit ATP6H (ATP6H) mRNA, complete cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

- 15 The disclosed NOV24 polypeptide (SEQ ID NO:58) encoded by SEQ ID NO:57 has 76 amino acid residues and is presented in Table 24B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV24 has a signal peptide and is likely to be localized in the plasma membrane with a certainty of 06400. The most likely cleavage site for a NOV24 peptide is between amino acids 53 and 54.

Table 24B. Encoded NOV24 protein sequence (SEQ ID NO:58).
<p>FALPVIIFTTFWGLVGIAGPWFVPKGPNRGVITMLVATAVCCYLFWLIAILAQLNPLFG PQLKNETIWIYVRFWE</p>

20

A search of sequence databases reveals that the NOV24 amino acid sequence has 56 of 73 amino acid residues (76%) identical to, and 64 of 73 amino acid residues (87%) similar to, the 81 amino acid residue ptnr:SPTREMBL-ACC:Q9N0Q1 protein from Canis familiaris

segments of ATP6F and ATP6C are 75% similar on the amino acid level. Northern blot analysis indicated that the 1.1-kb ATP6F mRNA was expressed in all tissues tested. The ATP6F gene contains 8 exons and spans approximately 4 kb. By FISH and radiation hybrid analysis, Nishigori et al. (1998) mapped the ATP6F gene to 1p32.3. (See Nishigori et al., Genomics 50: 222-228, 1998).

The vacuolar proton-ATPase (V-ATPase) is composed of an extramembrane catalytic sector and a transmembrane proton-conducting sector. See 603717. Ludwig et al. (1998) identified 2 novel proteins, 8-9 and 9.2 kD in size, in the membrane sector of bovine chromaffin granule V-ATPase. They designated the larger protein M9.2. By searching an EST database with the N-terminal sequence of bovine M9.2, Ludwig et al. (See Ludwig, et al., J. Biol. Chem. 273: 10939-10947, 1998) identified homologous cDNAs from human and mouse. The deduced 80-amino acid human M9.2 protein is extremely hydrophobic with 2 predicted membrane-spanning helices. Human and mouse M9.2 differed at only 1 amino acid position. Northern blot analysis revealed that M9.2 was present in all bovine tissues tested.

The disclosed NOV24 nucleic acid of the invention encoding a vacuolar proton-ATPase subunit H-like protein includes the nucleic acid whose sequence is provided in Table 24A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 24A while still encoding a protein that maintains its vacuolar proton-ATPase subunit H-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 27 percent of the bases may be so changed.

The disclosed NOV24 protein of the invention includes the vacuolar proton-ATPase subunit H-like protein whose sequence is provided in Table 24B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 24B while still encoding a protein that maintains its vacuolar proton-

ATPase subunit H-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 24 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

5 The above defined information for this invention suggests that this vacuolar proton-ATPase subunit H-like protein (NOV24) may function as a member of a “vacuolar proton-ATPase subunit H family”. Therefore, the NOV24 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this
10 invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

15 The NOV24 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the vacuolar proton-ATPase subunit H-like protein (NOV24) may be useful in gene therapy, and the vacuolar proton-ATPase subunit H-like protein (NOV24) may be useful when administered to a subject
20 in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from polycystic kidney disease I; osteopetrosis; mucopolipidosis IV, or other pathologies or conditions. The NOV24 nucleic acid encoding the vacuolar proton-ATPase subunit H-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of
25 the nucleic acid or the protein are to be assessed.

NOV24 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV24 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-
30 NOVX Antibodies” section below. The disclosed NOV24 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV25

A disclosed NOV25 nucleic acid of 5587 nucleotides (also referred to as CG57503-01) encoding a MEGF7-like protein is shown in Table 25A. The start and stop codons are in bold letters.

Table 25A. NOV25 nucleotide sequence (SEQ ID NO:59).
ATGGGCCTAGGAGTCATACTACCTACCTGTTCCCTCTTGACTTTCACTGTGACAATGGCAAGTGCATCC GCCGCTCCTGGGTGTGTGACGGGGACAACGACTGTGAGGATGACTCGGATGAGCAGGACTGTCCCCCCCG GGAGTGTGAGGAGGACGAGTTTCCCTGCCAGAATGGCTACTGCATCCGGAGTCTGTGGCACTGCGATGGT GACAATGACTGTGGCGACAACAGCGATGAGCAGTGTGACATGCGCAAGTGCTCCGACAAGGAGTTCCGCT GTAGTGACGGAAGCTGCATTTGCTGAGCATTGGTACTGCGACGGTGACACCGACTGCAAAGATGGCTCCGA TGAGGAGAACTGTCCCTCAGCAGTGCCAGCGCCCCCTGCAACCTGGAGGAGTTCCAGTGTGCCTATGGA CGCTGCATCCTCGACATCTACCACTGCGATGGCGACGATGACTGTGGAGACTGGTCAGACGAGTCTGACT GCTGTGAGTACTCTGGCCAGCTGGGAGCCTCCACCAGCCCTGCCGCTCTGGGGAGTTTATGTGTGACAG TGGCCTGTGTCATCAATGCAGGCTGGCGCTGCGATGGTGACGCGGACTGTGATGACCAGTCTGATGAGCGC AACTGCACCACCTCCATGTGTACGGCAGAACAGTTCCGCTGTCACTCAGGCCGCTGTGTCCGCTGTCTCT GGCGCTGTGATGGGGAGGACGACTGTGCAGACAACAGCGATGAAGAGAAGTGTGAGAATACAGGAAGCCC CCAATGTGCCTTGGACCAGTTCTGTGTGGAAATGGGCGCTGCATTGGGCAGAGGAAGCTGTGCAACGGG GTCAACGACTGTGGTGACAACAGCGACGAAAGCCACAGCAGAATTGCCGGCCCCGGACGGGTGAGGAGA ACTGCAATGTTAAACAACGGTGGCTGTGCCCAGAAGTGCCAGATGGTGCGGGGGGCGAGTGCAGTGTACCTG CCACACAGGCTACCGGCTCACAGAGGATGGGCACACGTGCCAAGATGTGAATGAATGTGCCGAGGAGGGG TATTGCGAGCCAGGGCTGCACCAACAGCGAAGGGGCTTTCCAATGCTGGTGTGAAACAGGCTATGAACACTAC GGCCCGACCGGCGCAGCTGCAAGGCTCTGGGGCCAGAGCCTGTGCTGCTGTTCCGAATCGCATCGACAT CCGGCAGGTGTGCCACACCGCTCTGAGTACACACTGCTGCTTAACAACCTGGAGAATGCCATTGCCCTT GATTTCCACCACCGCCGCGAGCTTGTCTTCTGGTTCAGATGTCAACCTGGACCGGATCTCCGTGCCAAC TCAACGGCAGCAACGCTGGAGGAGGTTGTGTCTACTGGGCTGGAGAGCCAGGGGGCTGGCTGTGGATTG GGTCATGACAACTCTACTGGACCGACTCAGGCACCTCGAGGATTGAGGTGGCAATCTGGACGGGGCC CACCGGAAAGTGTGTGTGTGGCAGAACCTGGAGAAGCCCCGGGCCATTGCCTTGATCCCATGGAGGGTA CCATTTACTGGACAGACTGGGGCAACCCCCGTATTGAGGCCTCCAGCATGGATGGCTCTGGACGCCG CATCATTTCCGATACCCATCTCTTCTGGCCCAATGGCCTCACCATCGACTATGCCGGGCGCGTATGTAC TGGGTGGATGCTAAGCACCATGTCTATCGAGAGGGCCAATCTGGATGGGAGTCAACCGTAAGGCTGTCAATTA GCCAGGGCCTCCCGCATCCCTTCGCCATCACAGTGTGTTGAAGACAGCCTGTACTGGACAGACTGGCACAC CAAGAGCATCAATAGCGCTAAACAAATTTACGGGGAAGAACCAGGAAATCATTGCAACAAACTCCACTTC CCTATGGACATCCACACCTTGCACCCCCAGCGCAACCTGCAGGGAAGAACCGCTGTGGGGACAACAACG GATCAGCAGCCACGCTGTGCCAGAGTCTTGACAAGTTCTGTCTTTTGGCCGAAGGATGGACATCCGT CGAATCAGCTTTGACACAGAGGACCTGTCTGATGATGTATCCACTGGCTGACGTGCGCAGTGTGTGG CCCTTGACTGGGACTCCCGGGATGACCACGTGTACTGGACAGATGTGACACTGTATACCATCAGCAGGGC CAAGTGGGATGGAACAGGACAGGAGGTGGTAGTGATACAGTTTGAGAGCCAGCTGGCCTGGCCATT GATTGGGTACCAACAAACTGTACTGGACAGATGCAAGGTACAGACCGGATTGAAGTAGCCAAACAGATG GCAGCATGAGAACAGTACTCATCTGGGAGAACCTTGATCGTCTCGGGACATCGTGGTGAACCCATGGG CGGTACATGTATTGGACTGACTGGGGTGCAGAGCCCAAGATTGAACGAGCTGGCATGGATGCCTCAGGC CGCCAAGTCATTATCTCTTAATCTGACCTGGCCTAATGGGTTAGCTATTGATTATGGGTCCCAGCGTC TATACTGGGCTGACCGCGCATGAAGACAATTGAATTTGTGGAAGTGGATGGCAGTAAGAGGAAGGTGCT GATTGGAAGCCAGCTCCCCACCCATTGGGCTGACCCTCTATGGAGAGCGCATCTATTGGACTGACTGG CAGACCAAGAGCATAAGAGCGCTGACCGGCTGACAGGGCTGGACCGGGAGACTCTGCAGGAGAACCTGG AAAACCTAATGGACATCCATGTCTTCCACCGCCCGCGGCCCCAGTGTCTACACCATGTGTATGGAGAA TGGCGGCTGTAGCCACCTGTGTCTTAGGTCCCCAAATCCAAGCGGATTGAGCTGTACCTGCCCCACAGGC ATCAACCTGTGTCTGATGGCAAGACCTGTCTACAGGATGAACAGTTTCTCATCTTCGCCAGGAGGA TAGACATTGCGATGGTCTCCCTGGACATCCCTTATTTTGTGTGATGTGGTGGTACCAATCAACATTACCAT GAAGAACACCATTTGCCATTGGAGTAGACCCCAAGGAAGGAAAGGTGACTGGTCTGACAGCACACTGCAC AGGATCAGTCGTGCCAATCTGGATGGCTCACAGCATGAGGACATCATCACCAAGGGCTACAGACCACAG ATGGGCTCGCGGTTGATGCCATTGGCCGGAAGTATGACTGGACAGACAGGGAACAAACCGGATTGAAGT GGGCAACCTGGACGGGTCCATGCGGAAAGTGTGGTGTGCGAGAACCTTGACAGTCCCGGGGCCATCGTA CTGTACCATGAGATGGGGTTTATGTACTGGACAGACTGGGGGAGAATGCCAAGTTAGAGCGGTCCGGAA TGGATGGCTCAGACCGCGCGGTGCTCATCAACAACAACCTAGGATGGCCCAATGGACTGACTGTGGACAA GGCCAGCTCCCAACTGCTATGGGCCGATGCCACACCGAGCGAATTGAGGCTGCTGACCTGAATGGTGCC AATCGGCATACATTGGTGTACCGGTGCAGCACCCATATGGCCTCACCTGTCTCGACTCTATCTACT GGACTGACTGGCAGACTCGGAGCATCCACCGTGTGACAAGGGTACTGGCAGCAATGTATCTCTGTGAG

```

GTCCAACCTGCCAGGCCTCATGGACATGCAGGCTGTGGACCGGGCACAGCCACTAGGTTTAAACAAGTGC
GGCTCGAGAAATGGCGGCTGCTCCCACTCTGCTTGGCTCGGCCTTCTGGCTTCTCTGTGCCTGCCCCA
CTGGCATCCAGCTGAAGGGAGATGGGAAGACCTGTGATCCCTCTCCTGAGACCTACCTGCTCTTCTCCAG
CCGTGGCTCCATCCGGCGTATCTCACTGGACACCAAGTGACACACCAATGTGCATGTCCCTGTTCTCTGAG
CTCAACAATGTCTCTCCCTGGACTATGACAGCGTGGATGGAAAGGTCTATTACACAGATGTGTTCTCTGG
ATGTTATCAGGCGAGCAGACCTGAACGGCAGCAACATGGAGACAGTGATCGGGCGAGGGCTGAAGACCAC
TGACGGGCTGGCAGTGGACTGGGTGGCCAGGAACCTGTACTGGACAGACACAGGTGAAATACCATTTAG
GCGTCCAGGCTGGATGGTTCTGCGCAAAGTACTGATCAACAATAGCCTGGATGAGCCCCGGGCCATTG
CTGTTTTCCCCAGGAAGGGGTACCTCTTCTGGACAGACTGGGGCCACATGCGCAAGATCGAACGGGCAAA
CTTGGATGGTTCTGAGCGGAAGGTCTCATCAACACAGACCTGGGTGGCCCAATGGCCTTACCCTGGAC
TATGATACCCGAGGATCTACTGGGTGGATGCGCATCTGGACCGGATCGAGAGTGTGACCTCAATGGGA
AACTGCGGCAGGTCTTGGTCAGCCATGTGTCCCAACCCCTTTGCCCTCACACAGCAAGACAGGTGGATCTA
CTGGACAGACTGGCAGACCAAGTCAATCCAGCGTGTGACAAATACTCAGGCCGAACAAGGAGACAGTG
CTGGCAAATGTGAAGGACTCATGGATATCATCGTGGTTTCCCCTCAGCGGCAGACAGGGACCAATGCCT
GTGGTGTGAACAATGGTGGCTGCACCCACCTCTGCTTTGCCAGAGCCTCGGACTTCGTATGTGCTGTGC
TGGAAACCTGATAGCCAGCCTGTCTCCCTTGTGCTGGCCTGGTACCACAGCTCCTAGGGTACTGGC
ATGAGTGAAAAGAGCCCAGTGCTACCCAACACACCACCTACCACCTTGTATTCTTCAACCACCCGGACCC
GCACGTCTCTGGAGGAGGTGAAGGAAGATGCTCTGAAAGGGATGCCAGGCTGGGCTCTGTGCACGTTT
CAATGACGCTGTTCTCTGTGCTCCAGGGGAAGGACTTCATATCAGTACGCCATTGGTGGACTCCTCAGT
ATTCTGCTGATTTTGGTGGTATTGACGCTTTGATGCTGTACAGACACAAAAATCCAAGTTCACTGATC
CTGGAATGGGGAACCTCACTACAGCAACCCCTCTACCGAACATCCACACAGGAAGTGAAGATTGAAGC
AATCCCCAAACCAGCCATGTACAACCAGCTGTGCTATAAGAAAGAGGGAGGGCCTGACCATAACTACACC
AAGGAGAAGATCAAGATCGTAGAGGAATCTGCCTCCTGTCTGGGGATGATGCTGAGTGGGATGACCTCA
AGCAACTGCGAAGCTCACGGGGGGGCTCCTCCGGGATCATGTATGCATGAAGACAGACAGGTGTCCAT
CCAGGCAGCTCTGGCTCCTGGATGACACAGAGCGGAGCAGCTGTTACAGGAAGAGCAGTCTGAGTGT
AGCAGCGTCCATACTGCAGCCACTCCAGAAAGACGAGGCTCTCTGCCAGACACGGGCTGGAACATGAAC
GCAAGCTCTCCTCAGAGAGCCAGGTCTAAATGCCACATTCTCTTCCCTGCCTGCCT

```

In a search of public sequence databases, the NOV25 nucleic acid sequence, located on chromosome 11 has 4754 of 4759 bases (99%) identical to a gb:GENBANK-

ID:AB011540|acc:AB011540.1 mRNA from Homo sapiens (Homo sapiens mRNA for

- 5 MEGF7, partial cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV25 polypeptide (SEQ ID NO:60) encoded by SEQ ID NO:59 has 1852 amino acid residues and is presented in Table 25B using the one-letter amino acid code.

Signal P, Psort and/or Hydropathy results predict that NOV25 has no signal peptide and is

- 10 likely to be localized at the plasma membrane with a certainty of 0.8200.

Table 25B. Encoded NOV25 protein sequence (SEQ ID NO:60).

```

MGLGVILPTCSPLDFHCDNGKCIIRSWVCDGDNDCEDDSDEQDCPPRECEDEFPQNGY
CIRSLWHCDGDNDCGDNDSDEQCDMRKCSDEKFRCSGSCIAEHWYCDGDTDCCKDGSDEEN
CPSAVPAPPNLEEFQCA YGRCILDIYHCDGDDCGDWSDESDCCEYSGQLGASHQPCRS
GEFMCDSGLCINAGWRCDGDADDDQSDERNCTTSMCTAEQFRCHSGRCVRLSWRCDGED
DCADNSDEENCENTGSPQCALDQFLCWNGRCIGQRKLCNGVNDCGDNDESPPQNCRPRT
GEENCNVNNGGCAQKQCMVRGAVQCTCHTGYRLTEDGHTCQDVNECAEEGYCSQGCTNSE
GAFQCWCETGYELRPDRRSCKALGPEPVLLFANRIDIRQVLPHRSEYTLNLENALIAL
DFHHRRELVFWSVDVTLDRILRANLNGSNVEEVVSTGLSPGGLAVDWVHDKLYWTDSGTS
RIEVANLDGAHRKVLLWQNLKPRALHHPMEGTIYWTDWGNTPRIEASSMDGSGRRIIA
DTHLFWPNGLTIDYAGRRMYWVDAKHVIERANLDGSHRKAVISQGLPHPFATVTFEDSL
YWTDWHTKSINSANKFTGKNQEIIRNKLHFPMDIHTLHPQRQPAGKNRCGDNNGGCTHLC
LPSGQNYTCACPTGFRKISSHACAQSLDKFLFARRMDIRRISFDTELDSDVIPLADVR
SAVALDWDSRDDHVYWTDVSTDTISRKWDGTGQEVVVDTSLES PAGLAIDWVTNKLYWT
DAGTDRIEVANTDGSMTVLWIENLDRPRDIVVEPMGGYMYWTDWGASPKIERAGMDASG
RQVIISNLTWPNGLAIDYGSQRLYWADAGMKTEIFAGLDGSKRKVLIGSQLPHPFGLTL
YGERIYWTDWQTKSIQSADRLTGLDRETLQENLENLMDIHVFHRRRPPVSTPCAMENGGC

```

VPINITMKNTIAIGVDPQEGKVYWSDSLHRISRANLDGSQHEDIITTGLQTTDGLAVDA
 IGRKVYWTDTGTNRIEVGNLDGSMRKVLVWQNLDSRAIVLYHEMGFMYWTDWGENAKLE
 RSGMDGSDRAVLINNGLWPNGLTVDKASSQLLWADAHTERIEAADLNGANRHTLVSPVQ
 HPYGLTLLDSYIYWTDWQTRSIRHADKGTGSNVILVRSNLPGLMDMQAVDRAQPLGFNKC
 GSRNGGCSHLCLPRPSGFSCACPTGIQLKGDGKTCDSPETYLFLSSRGSIRRISLDTSD
 HTNVHVPVPELNNVISLDYDSVDGKVYYTDFLDVIRRADLNGSNMETVIGRGLKTTDGL
 AVDWVARNLYWTDGTGRNTIEASRLDGSCRKVLINNSLDEPRAIAVFPKGYLFWTDWGH
 AKIERANLDGSEKVLINTDLGWPNGLTLDYDTRRIYWVDAHLDRIESADLNGKLRQVLV
 SHVSHFPALTQQDRWIYWTDWQTKSIQVRDKYSGRNKETVLANVEGLMDIIVSPQRQTG
 TNACGVNNGGCTHLCFARASDFVCACPDEPDSQPCSLVPGLVPPAPRATGMSEKSPVLN
 TPPTTLYSSTTRTRTSLEEVEGRCSERDARLGLCARSNDAPPAAPGEGHLISYAIGGLLS
 ILLILVVIAALMLYRHKKSFTDPGMGNLTYSNPSYRTSTQEVKIEAIPKPMYNQLCYK
 KEGGPDHNYTKEKIKIVEGICLLSGDDAEWDDLKQLRSSRGGLLRDHVCMKTDTVSIQAS
 SGLSDDTETEQLLQEEQSECSSVHTAATPERRGSLPDTGWKHERKLSSSESQV

A search of sequence databases reveals that the NOV25 amino acid sequence has 1572 of 1576 amino acid residues (99%) identical to, and 1574 of 1576 amino acid residues (99%) similar to, the 1576 amino acid residue ptnr:SPTREMBL-ACC:O75096 protein from Homo sapiens (Human) (MEGF7). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV25 is expressed in at least adrenal gland/suprarenal gland, bone marrow, brain, bronchus, brown adipose, cartilage, cervix, colon, heart, hypothalamus, lung, peripheral blood, pituitary gland, spinal chord, stomach, testis, thalamus, uterus. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV25 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 25C.

15

Table 25C. BLAST results for NOV25					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17224416 gb AAL36970.1 (AF247637)	LDLR dan [Mus musculus]	1905	1779/1849 (96%)	1809/1849 (97%)	0.0
gi 3449306 dbj BAA32468.1 (AB011540)	MEGF7 [Homo sapiens]	1576	1572/1576 (99%)	1574/1576 (99%)	0.0
gi 6681362 dbj BAA88688.1 (AB011533)	MEGF7 [Rattus norvegicus]	1298	1248/1298 (96%)	1274/1298 (98%)	0.0
gi 17472590 ref XP_061753.1 (XM_061753)	similar to MEGF7 (H. sapiens)	1007	992/992 (100%)	992/992 (100%)	0.0

gi 14763921 ref XP_035037.1 (XM_035037)	low density lipoprotein receptor-related protein 4 [Homo sapiens]	859	857/859 (99%)	859/859 (99%)	0.0
---	---	-----	---------------	---------------	-----

Tables 25D-E list the domain descriptions from DOMAIN analysis results against NOV25. This indicates that the NOV25 sequence has properties similar to those of other proteins known to contain this domain.

Table 25D. Domain Analysis of NOV25

gnl|Smart|smart00192, LDLa, Low-density lipoprotein receptor domain class A; Cysteine-rich repeat in the low-density lipoprotein (LDL) receptor that plays a central role in mammalian cholesterol metabolism. The N-terminal type A repeats in LDL receptor bind the lipoproteins. Other homologous domains occur in related receptors, including the very low-density lipoprotein receptor and the LDL receptor-related protein/alpha 2-macroglobulin receptor, and in proteins which are functionally unrelated, such as the C9 component of complement. Mutations in the LDL receptor gene cause familial hypercholesterolemia.

CD-Length = 38 residues, 97.4% aligned

Score = 64.3 bits (155), Expect = 6e-11

Table 25E. Domain Analysis of NOV25

gnl|Smart|smart00135, LY, Low-density lipoprotein-receptor YWTD domain; Type "B" repeats in low-density lipoprotein (LDL) receptor that plays a central role in mammalian cholesterol metabolism. Also present in a variety of molecules similar to gp300/megalin.

CD-Length = 43 residues, 95.3% aligned

Score = 62.0 bits (149), Expect = 3e-10

The domain that characterizes epidermal growth factor (EGF) consists of approximately 50 amino acids, and has been shown to be present, in a more or less conserved form, in a large number of other, mostly animal proteins. EGF-like domains are believed to play a critical role in a number of extracellular events, including cell adhesion and receptor-ligand interactions. Proteins with EGF-like domains often consist of more than 1,000 amino acids, have multiple copies of the EGF-like domain, and contain additional domains known to be involved in specific protein-protein interactions. The list of proteins currently known to contain one or more copies of an EGF-like pattern is large and varied. The functional significance of EGF domains in what appear to be unrelated proteins is not yet clear. However,

a common feature is that these repeats are found in the extracellular domain of membrane-bound proteins or in proteins known to be secreted (exception: prostaglandin G/H synthase). The EGF domain includes six cysteine residues which have been shown (in EGF) to be involved in 3 disulfide bonds. The main structure is a two-stranded beta-sheet followed by a loop to a C-terminal short two-stranded sheet. Subdomains between the conserved cysteines vary in length.

To identify proteins containing EGF-like domains, Nakayama et al. (1998) searched a database of long cDNA sequences randomly selected from a human brain cDNA library for those that encode an EGF-like motif. They identified several partial cDNAs encoding novel proteins with EGF-like domains, such as LRP4, which they named MEGF7. The predicted partial LRP4 protein contains 2 EGF-like domains, a calcium binding-type EGF-like domain, 3 LDL receptor-type EGF-like domains, 4 YWTD spacer regions, a transmembrane domain, a cytoplasmic NPXY motif, which is required for clustering and internalization of LDL receptors, and a cytoplasmic tSXV motif, which anchors proteins with a PDZ domain. The sequence and domain organization of LRP4 shows significant similarities to those of members of the LDL receptor family. Northern blot analysis detected rat Megf7 expression in several regions of the brain. Using a radiation hybrid mapping panel, Nakayama et al. (1998) mapped the LRP4 gene to 11p12-p11.2.

The disclosed NOV25 nucleic acid of the invention encoding a MEGF7-like protein includes the nucleic acid whose sequence is provided in Table 25A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 25A while still encoding a protein that maintains its MEGF7-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 1 percent of the bases may be so changed.

The disclosed NOV25 protein of the invention includes the MEGF7-like protein whose sequence is provided in Table 25B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 25B while still encoding a protein that maintains its MEGF7-like activities and physiological
5 functions, or a functional fragment thereof. In the mutant or variant protein, up to about 1 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this MEGF7-like
10 protein (NOV25) may function as a member of a "MEGF7 family". Therefore, the NOV25 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein
15 therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV25 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies
20 and disorders as indicated below. For example, a cDNA encoding the MEGF7-like protein (NOV25) may be useful in gene therapy, and the MEGF7-like protein (NOV25) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from adrenoleukodystrophy, congenital adrenal hyperplasia, hemophilia, hypercoagulation,
25 idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, graft vesus host; diseases of the brain and nervous system, including Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral
30 disorders, addiction, anxiety, pain, neuroprotection; diseases of the respiratory system, including systemic lupus erythematosus , autoimmune disease, asthma, emphysema, scleroderma, allergy, ARDS; diseases and disorders of adipose tissue, reproductive system, colon, circulatory system, spinal chord, digestive system, and endocrine system, or other pathologies or conditions. The NOV25 nucleic acid encoding the MEGF7-like protein of the

invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV25 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV25 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV25 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV26

A disclosed NOV26 nucleic acid of 635 nucleotides (also referred to as CG57456-01) encoding a COP-Coated Vesicle Membrane Protein P24 Precursor-like protein is shown in Table 26A. The start and stop codons are in bold letters.

Table 26A. NOV26 nucleotide sequence (SEQ ID NO:61).

```
CCCCACTATGGTGACGCTCGCTGAGCTGCTGTTGCTCCAGAACACTCTCCTGACCATGGTCTTGGGCTAT
TTCATCAGCATCCACGCACATGCTGAAGAATGCTTAAGTGAGCATGTCACCTCAGGCACCAAGATGGGCC
TCATCTTCTGAAGGTGGCTTCCTGGGCATCAACATGGAGATTACAGGACCTAAGAATAAAGGATTATATAA
AGGAGACAAAGAATCCAGTGGGAAATACACATTTCTGCTCAGATGGATGGAACAAATACATTTTGTGTTT
AGTGACCGAGTGTCCACCATGACTCCAAAGATAGTGATATTCACCATTGATATTGGGGAGGCTACAAAAA
GAGAAGACATGGAACAGAAGCTCACCAGAACAACTAGAAGAAATGATCAGTGAGCTGGCTGTGGCCAT
GACAGCTGTACAGCACAAGAGGAATACACGAAAATCTGGGAGAGGATACACAGAGCCATTAGTGACAAC
ACAAACAGCCCAGTGGTCCTTCGGTGCTTCTTTGAAGCTCTTGTTCTAATTGCCATGACATTGGGACACA
TCTACTACCTGAAGAGATTTTTTGAAGTCCAGAGGGTTGTTTCAAAGCCTCTTCCTGATGATTCCAAC
TCATA
```

In a search of public sequence databases, the NOV26 nucleic acid sequence has 630 of 635 bases (99%) identical to a gb:GENBANK-ID:AF152363|acc:AF152363.1 mRNA from Homo sapiens (Homo sapiens constitutive fragile region FRA3B sequence). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV26 polypeptide (SEQ ID NO:62) encoded by SEQ ID NO:61 has 203 amino acid residues and is presented in Table 26B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV26 has a signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.4600. The most likely cleavage site for a NOV26 peptide is between amino acids 29 and 30.



Table 26B. Encoded NOV26 protein sequence (SEQ ID NO:62).
MVTLAELLLLQNTLLTMVLGYFISIHAAEECLSEHVTSGTKMGLIFEGGFLGINMEITG PKNKRIYKGDKESSGKYTFSAHMDGTNTFCFSDRVSTMTPKIVIFTIDIGEATKREDMET EAHQNKLEEMISELAVAMTAVQHKEEYTKIWERIHRAISDNTNSPVVLRCCFFEALVLIAM TLGHIYYLKRFFEVRVSKASS

A search of sequence databases reveals that the NOV26 amino acid sequence has 156
 of 201 amino acid residues (77%) identical to, and 175 of 201 amino acid residues (87%)
 similar to, the 201 amino acid residue ptmr:SWISSNEW-ACC:Q15363 protein from Homo
 sapiens (Human) (COP-COATED VESICLE MEMBRANE PROTEIN P24 PRECURSOR
 (P24A) (RNP24)). Public amino acid databases include the GenBank databases, SwissProt,
 PDB and PIR.

NOV26 is expressed in at least Eye, placenta, colon, and ovary. This information was
 derived by determining the tissue sources of the sequences that were included in the invention
 including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or
 RACE sources.

The disclosed NOV26 polypeptide has homology to the amino acid sequences shown
 in the BLASTP data listed in Table 26C.

Table 26C. BLAST results for NOV26					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17440558 ref XP_067504.1 (XM_067504)	similar to coated vesicle membrane protein [Homo sapiens]	271	186/192 (96%)	186/192 (96%)	e-100
gi 9790015 ref NP_062744.1 (NM_019770)	coated vesicle membrane protein; Sid394p [Mus musculus]	201	155/201 (77%)	173/201 (85%)	4-74
gi 1352660 sp P49020 P24_CRIGR	Cop-coated vesicle membrane protein p24 precursor	196	148/196 (75%)	169/196 (85%)	6-74
gi 5803149 ref NP_006806.1 (NM_006815)	coated vesicle membrane protein [Homo sapiens]	201	156/201 (77%)	175/201 (86%)	e-73
gi 13929014 ref NP_113910.1 (NM_031722)	coated vesicle membrane protein [Rattus norvegicus]	201	155/201 (77%)	174/201 (86%)	e-73

Table 26D lists the domain descriptions from DOMAIN analysis results against NOV26. This indicates that the NOV26 sequence has properties similar to those of other proteins known to contain this domain.

Table 26D. Domain Analysis of NOV26

gnl|Pfam|pfam01105, EMP24_GP25L, emp24/gp25L/p24 family. Members of this family are implicated in bringing cargo forward from the ER and binding to coat proteins by their cytoplasmic domains.

CD-Length = 202 residues, 92.6% aligned

Score = 135 bits (341), Expect = 2e-33

Members of the p24 family of putative cargo receptors are proposed to contain retrograde and anterograde trafficking signals in their cytoplasmic domain to facilitate coat protein binding and cycling in the secretory pathway. The localization and transit of the wild-type chimera from the endoplasmic reticulum (ER) through the Golgi complex involved a glutamic acid residue and a conserved glutamine in the TMD. The TMD glutamic acid mediated the localization of the chimeras to the ER in the absence of the conserved glutamine. Efficient ER exit required the TMD glutamine and was further facilitated by a pair of phenylalanine residues in the cytoplasmic tail. TMD residues of p24 proteins may mediate the interaction with integral membrane proteins of the vesicle budding machinery to ensure p24 packaging into transport vesicles.

Blum et al. (1996) identified a 21-kD rat pancreatic microsomal membrane protein that they designated Tmp21. By probing a human brain cDNA library with a fragment of the rat sequence, they isolated a cDNA encoding human TMP21. The deduced 219-amino acid type I intracellular transmembrane protein contains a signal sequence and is predicted to be located in the lumen of the endoplasmic reticulum. Northern blot analysis detected a 1.4-kb TMP21 transcript. Immunoblot analysis showed that the rat Tmp21 protein is expressed predominantly in the microsomal fraction of pancreatic acinar cells. Horer et al. (1999) determined that a putative TMP21 isoform, TMP21-II, is a neutral pseudogene.

The disclosed NOV26 nucleic acid of the invention encoding a COP-Coated Vesicle Membrane Protein P24 Precursor-like protein includes the nucleic acid whose sequence is provided in Table 26A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 26A while still encoding a protein that maintains its COP-Coated Vesicle Membrane Protein P24 Precursor-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or

complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 1 percent of the bases may be so changed.

The disclosed NOV26 protein of the invention includes the COP-Coated Vesicle Membrane Protein P24 Precursor-like protein whose sequence is provided in Table 26B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 26B while still encoding a protein that maintains its COP-Coated Vesicle Membrane Protein P24 Precursor-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 23 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this COP-Coated Vesicle Membrane Protein P24 Precursor-like protein (NOV26) may function as a member of a "COP-Coated Vesicle Membrane Protein P24 Precursor family". Therefore, the NOV26 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV26 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the COP-Coated Vesicle Membrane Protein P24 Precursor-like protein (NOV26) may be useful in gene therapy, and the COP-Coated Vesicle Membrane Protein P24 Precursor-like protein (NOV26) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from Endometriosis, Fertility, Von Hippel-Lindau (VHL) syndrome, Diabetes, Tuberous

sclerosis, or other pathologies or conditions. The NOV26 nucleic acid encoding the COP-Coated Vesicle Membrane Protein P24 Precursor-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV26 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV26 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV16 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV27

A disclosed NOV27 nucleic acid of 1120 nucleotides (also referred to as CG57658-01) encoding a connexin-like protein is shown in Table 27A. The start and stop codons are in bold letters.

Table 27A. NOV27 nucleotide sequence (SEQ ID NO:63).

GAGGCCATGCCCGCTTCTCTCTTCCAGGAAAGCTCTGGTTCGTCTCACGATGCTGCTGCGGATGCTGG
TGATTGCTTTGGCGGGGCGACCCGTCTACCAGGACGAGCAGGAGAGGTTTGTCTGCAACACGCTGCAGCC
GGGATGCGCCAATGTTTGCTACGACGTCTTCTCCCCGTGTCTCACCTGCGGTTCTGGCTGATCCAGGGC
GTGTGCGTCTCTCTCCCTCCGCGGTCTTCAGCGTCTATGTCCTGCACCGAGGAGCCACGCTCGCCGCGC
TGGGCCCCCGCGCTGCCCCGACCCCCGGGAGCCGGCTCCGGGCAGAGACGCTGCCCCGCGCCATTTCGG
GGAGCGCGGCGGCCTCCAGGTGCCCCGACTTTTCGGCCGGCTACATCATCCACCTCCTCCTCCGGACCCTG
CTGGAGGCAGCCTTCGGGGCCTTGCACTACTTTCTCTTTGGATTCTTGCCCCGAAGAAGTTCCTTGCA
CGCGCCCTCCGTGCACGGGCGTGGTGGACTGCTACGTGTGCGGCCCCACAGAGAAGTCCCTGCTGATGCT
GTTCTCTTGGGCGGTGAGCGCGTGTCTTTCTGCTGGGCCTCGCCGACCTGGTCTGCAGCCTGCGGCGG
CGGATGCGCAGGAGGCCGGGACCCCCACAAGCCCTCCATCCGGAAGCAGAGCGGAGCCTCAGGCCACG
CGGAGGGACGCCGACTGACGAGGAGGTGGGCGGGAGGAAGAGGGGGCACCGGCGCCCCCGGGTGACG
CGCCGAGGGGAGGGGGCTGGCAGCCCCAGGCGTACATCCAGGGTGTGAGGGCACACGAAGATTCCGGAT
GAGGATGAGAGTGAGGTGACATCCTCCGCCAGCGAAAAGCTGGGCAGACAGCCCCGGGGCAGGCCCCACC
GAGAGGCCGCCAGGACCCAGGGGCTCAGGATCCGAGGAGCAGCCCTCAGCAGCCCCCAGCCGCTGGC
CGCGCCCCCTTCTGTCAGCAGCCTGCAGCCCCCTGACCCGCTGCCAGCTCCAGTGGTGTCCCCACCTG
AGAGCCAGGAAGTCTGAGTGGGTGTGAAAAAACAGCACCTGGCGGTGCCCCGGGGCTCACGCCTGTAAT

In a search of public sequence databases, the NOV27 nucleic acid sequence, located on chromosome 10 has 1037 of 1097 bases (94%) identical to a gb:GENBANK-ID:AB046017|acc:AB046017.1 mRNA from Macaca fascicularis (Macaca fascicularis brain

cDNA, clone:QccE-15512). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

- The disclosed NOV27 polypeptide (SEQ ID NO:64) encoded by SEQ ID NO:63 has 356 amino acid residues and is presented in Table 27B using the one-letter amino acid code.
- 5 Signal P, Psort and/or Hydropathy results predict that NOV27 has a signal peptide and is likely to be localized in the plasma membranewith a certainty of 0.6400. The most likely cleavage site for a NOV27 peptide is between amino acids 26 and 27 .

Table 27B. Encoded NOV27 protein sequence (SEQ ID NO:64).	
MPASSLPGKLWFLVTMLLRMLVIVLAGRPVYQDEQERFVCNTLQPGCANVCYDVFSPVSH LRFWLIQGVCVLLPSAVFSVYVLRHGATLAALGPRRCPPREPASGQRRCPRPFGERGGL QVPDFSAGYI IHLLLRRTLLEAAGALHYFLFGFLAPKKFPCTRPCTGVVDCYVSRPTEK SLLMLFLWAVSALSFLGLADLVCSLRRRMRRRPGPPTSPSIRKQSGASGHAEGRRTDEE GGREEEGAPAPPGARAGGEGAGSPRRTSRVSGHTKIPDEDESEVTSSASEKLGRQPRGRP HREAAQDPRGSGSEEQPSAAPSRLAAPPSCSSLQPPDPASSSGAPHLRARKSEW	

- A search of sequence databases reveals that the NOV27 amino acid sequence has
- 10 348/348 (100%) identical to TREMBLNEW-ACC:CAC10186 BA425A6.2 (SIMILAR TO CONNEXIN) - Homo sapiens. Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

- NOV27 is expressed in at least Brain, Lung, Ovary, colon. This information was derived by determining the tissue sources of the sequences that were included in the invention
- 15 including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV27 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 27C.

Table 27C. BLAST results for NOV27					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17489790 ref XP_058368.1 (XM_058368)	similar to bA425A6.2 (similar to connexin) [Homo sapiens]	370	349/352 (99%)	351/352 (99%)	e-167
gi 10334641 emb CAC10186.1 (AL121749)	bA425A6.2 (similar to connexin) [Homo sapiens]	348	348/348 (100%)	348/348 (100%)	e-163
gi 9280090 dbj BAB01599.1 (AB046017)	unnamed protein product [Macaca fascicularis]	341	316/341 (92%)	323/341 (94%)	e-159
gi 17489782 ref XP_061277.1 (XM_061277)	similar to connexin) [Homo sapiens]	341	341/341 (100%)	341/341 (100%)	e-158

gi 15990849 emb CAC 93844.1	(AJ414562)	connexin39 [Mus musculus]	364	175/372 (47%)	215/372 (57%),	6e-67
--------------------------------	------------	------------------------------	-----	------------------	-------------------	-------

Tables 27D-E list the domain descriptions from DOMAIN analysis results against NOV27. This indicates that the NOV27 sequence has properties similar to those of other proteins known to contain this domain.

Table 27D. Domain Analysis of NOV27

gnl|Pfam|pfam00029, connexin, Connexin.

CD-Length = 218 residues, 91.7% aligned

Score = 172 bits (436), Expect = 3e-44

Table 27E. Domain Analysis of NOV27

gnl|Smart|smart00037, CNX, Connexin homologues; Connexin channels participate in the regulation of signaling between developing and differentiated cell types.

CD-Length = 34 residues, 97.1% aligned

Score = 64.7 bits (156), Expect = 9e-12

Gap junctions were first characterized by electron microscopy as regionally specialized structures on plasma membranes of contacting adherent cells. These structures were shown to consist of cell-to-cell channels. Proteins, called connexins, purified from fractions of enriched gap junctions from different tissues differ. The connexins are designated by their molecular mass. Another system of nomenclature divides gap junction proteins into 2 categories, alpha and beta, according to sequence similarities at the nucleotide and amino acid levels. For example, CX43 is designated alpha-1 gap junction protein, whereas CX32 and CX26 are called beta-1 and beta-2 gap junction proteins, respectively. This nomenclature emphasizes that CX32 and CX26 are more homologous to each other than either of them is to CX43. The connexins are a family of integral membrane proteins that oligomerise to form intercellular channels that are clustered at gap junctions. These channels are specialised sites of cell-cell contact that allow the passage of ions, intracellular metabolites and messenger molecules (with molecular weight <1-2 kD) from the cytoplasm of one cell to its apposing neighbours. They are found in almost all vertebrate cell types, and somewhat similar proteins have been cloned from plant species. Invertebrates utilise a different family of molecules, innexins, that share a similar predicted secondary structure to the vertebrate connexins, but have no sequence identity to them. Vertebrate gap junction channels are thought to participate in diverse

biological functions. For instance, in the heart they permit the rapid cell-cell transfer of action potentials, ensuring coordinated contraction of the cardiomyocytes. They are also responsible for neurotransmission at specialised 'electrical' synapses. In non-excitabile tissues, such as the liver, they may allow metabolic cooperation between cells. In the brain, glial cells are

5 extensively-coupled by gap junctions; this allows waves of intracellular Ca^{2+} to propagate through nervous tissue, and may contribute to their ability to spatially-buffer local changes in extracellular K^{+} concentration. The connexin protein family is encoded by at least 13 genes in rodents, with many homologues cloned from other species. They show overlapping tissue expression patterns, most tissues expressing more than one connexin type. Their conductances,

10 permeability to different molecules, phosphorylation and voltage-dependence of their gating, have been found to vary. Possible communication diversity is increased further by the fact that gap junctions may be formed by the association of different connexin isoforms from apposing cells. However, in vitro studies have shown that not all possible combinations of connexins produce active channels. Hydropathy analysis predicts that all cloned connexins share a

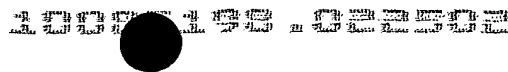
15 common transmembrane (TM) topology. Each connexin is thought to contain 4 TM domains, with two extracellular and three cytoplasmic regions. This model has been validated for several of the family members by in vitro biochemical analysis. Both N- and C-termini are thought to face the cytoplasm, and the third TM domain has an amphipathic character, suggesting that it contributes to the lining of the formed-channel. Amino acid sequence

20 identity between the isoforms is ~50-80%, with the TM domains being well conserved. Both extracellular loops contain characteristically conserved cysteine residues, which likely form intramolecular disulphide bonds. By contrast, the single putative intracellular loop (between TM domains 2 and 3) and the cytoplasmic C-terminus are highly variable among the family members. Six connexins are thought to associate to form a hemi-channel, or connexon. Two

25 connexons then interact (likely via the extracellular loops of their connexins) to form the complete gap junction channel.

The disclosed NOV27 nucleic acid of the invention encoding a connexin-like protein includes the nucleic acid whose sequence is provided in Table 27A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed

30 from the corresponding base shown in Table 27A while still encoding a protein that maintains its connexin-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or



complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 6 percent of the bases may be so changed.

The disclosed NOV27 protein of the invention includes the connexin-like protein whose sequence is provided in Table 27B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table B while still encoding a protein that maintains its connexin-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 0 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this connexin-like protein (NOV27) may function as a member of a “connexin family”. Therefore, the NOV27 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV27 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the connexin-like protein (NOV27) may be useful in gene therapy, and the connexin-like protein (NOV27) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from Cardiomyopathy, Atherosclerosis, Hypertension, Congenital heart defects, Aortic stenosis, Atrial septal defect (ASD), Atrioventricular (A-V) canal defect, Ductus arteriosus, Pulmonary stenosis, Subaortic stenosis, Ventricular septal defect (VSD), valve diseases, Tuberosus sclerosis, Scleroderma, Obesity, Transplantation, Diabetes, Von Hippel-Lindau (VHL)

syndrome , Pancreatitis,Obesity, Endometriosis,Fertility, Hemophilia,
Hypercoagulation,Idiopathic thrombocytopenic purpura , Immunodeficiencies,Graft vesus
host, Autoimmune disease, Renal artery stenosis, Interstitial nephritis, Glomerulonephritis,
Polycystic kidney disease, Systemic lupus erythematosus, Renal tubular acidosis, IgA
5 nephropathy, Hypercalceimia, Lesch-Nyhan syndrome, Von Hippel-Lindau (VHL) syndrome ,
Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalceimia, Parkinson's disease,
Huntington's disease, Cerebral palsy, Epilepsy,Lesch-Nyhan syndrome, Multiple
sclerosis,Ataxia-telangiectasia,Leukodystrophies,Behavioral disorders, Addiction, Anxiety,
Pain, Neuroprotection, or other pathologies or conditions. The NOV27 nucleic acid encoding
10 the connexin-like protein of the invention, or fragments thereof, may further be useful in
diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are
to be assessed.

NOV27 nucleic acids and polypeptides are further useful in the generation of
antibodies that bind immuno-specifically to the novel NOV27 substances for use in
15 therapeutic or diagnostic methods. These antibodies may be generated according to methods
known in the art, using prediction from hydrophobicity charts, as described in the “Anti-
NOVX Antibodies” section below. The disclosed NOV27 proteins have multiple hydrophilic
regions, each of which can be used as an immunogen. These novel proteins can be used in
assay systems for functional analysis of various human disorders, which will help in
20 understanding of pathology of the disease and development of new drug targets for various
disorders.

NOV28

A disclosed NOV28 nucleic acid of 1234 nucleotides (also referred to as CG57662-01)
encoding a -like protein is shown in Table 28A. The start and stop codons are in bold letters.

25

Table 28A. NOV28 nucleotide sequence (SEQ ID NO:65).

<p>TAATAATCTTTTAACTCCCTAACAGGATGTGTGGCAGGTTCTGAGGTGGTGGCTGCTGGCGGAG GAGAGCTGGCACTCCACCCCGTGGGGCGCTCTGTTTCCCGTGCTCCTGGGATTCCGCCTTGTGCTGC TGGCTGCCAGTGGGCCTGGAGTCTATGGCGATGAGCAGAGTGAATTCGTGTGTACACCCAGCAGCCGGG CTGCAAGGCTGCTGCTTCGATGCCTTCCACCGCTCTCCCGCTGCGTTTCTGGGTCTTCCAGGTCATC TTGGTGGCTGTACCTAGCGTCTCTACATGGGTTTCACTCTGTATCACGTGATCTGGCACTGGGAAGAAT CAAGAAAGGGGACGGAGGAAGAGGACACCCGATCCAGGGAGGGGAGAGCAGCAGAGATACCCAGGGGC TGGAAGCCTCAGGCTGCTCCGAGCTTATGTGGCTCAGCTGGGAGCTCAGCTGGTCTTGGAGGGGACAGCG CCGGGTTGTCAGTACCACCTGTATGGGTTCCAGATGCCAGCTCCTTTGCATGTGGCCAAAGACCTTGCC CGTATAGATTAACTTGCACCTTTTCCACCCCTCGGAGAAGATCATCTTTCTAAAAGCCATGTTTGGGGT CAGTGGGTTCCGTCTCTTGTCACTCTTTTGGAGATTGTGCTTCTGGGTCTGGGAAGACTGTGTAAGCCC CTGCGGAACCTTCTGGGTGGGGCTCTTCTCCAGCCACGCCCTGGCCCTGAGCAGCAAAAGGAACCTCC AGCAGACACTGGGAGCCATCCATCGGCCTGGTCAGCCTTGTTCATTTCAGAGACCATGTTCCCCACAGC CCAGTGTACTAGGGGTGACATCTCCCGACCTCCCCACCTGTGGATATGGCCAAGTCGAGGTACCGGTTA ACCAAAGATGCTGAAGGAGTGAAGAACCAGCCATCCCCTAATACGCAGGATGGTTATATTGATTATGTCA</p>

AACTGAAAACCTTTGGAGAACTCCTCTCTCAGAAAGCGATAACTGGGCCAGACACGGTGGCTCATGCCTG
 TAATCCCAGCATTTTGGGAGGCCTAGGCAGGTGATCACTGGAGGTGAGGAGTTCAAGACCAGCCAGGCC
 AACATGGTGAAACCCGTGTCTACTAAACTACAAAAATTCTGGGCATGGTGGTGGGCGTCTGTAATCCCA
 GCTACTTGAGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAG

In a search of public sequence databases, the NOV28 nucleic acid sequence, located on chromosome 7 has 206 of 244 bases (84%) identical to a gb:GENBANK-ID:AP000692|acc:AP000692.1 mRNA from Homo sapiens (Homo sapiens genomic DNA, chromosome 21q22.2, PAC clone:24J14, CBR1-HLCS region). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV28 polypeptide (SEQ ID NO:66) encoded by SEQ ID NO:65 has 391 amino acid residues and is presented in Table 28B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV28 has a signal peptide and is likely to be localized in the plasma membrane with a certainty of 0.6000. The most likely cleavage site for a NOV28 peptide is between amino acids 46 and 47.

Table 28B. Encoded NOV28 protein sequence (SEQ ID NO:66).

MCGRFLRWLLAEESWHSTPVGRLLPVLLGFRLVLLAASGPGVYGDEQSEFVCHTQQPG
 CKAACFDAFHPLSPLRFVVFQVILVAVPSVLYMGFTLYHVIWHWEESRKGTEEDTLIQG
 GESSRDTPGAGSLRLLRAYVAQLGAQLVLEGTAPGLQYHLYGFQMPSSFACGQEPYRL
 TCTFSPSEKIIIFLKAMFGVSGFRLLFTLLEIVLLGLGRCLKPLRNFLGGASSSSHALAL
 SSKRNLQQTLAGAIHRPGQPCSISETMFPTAPVTRGDISRPPPPVDMAKSRYRLTKDAEGV
 KNQSPSPNTQDGYIDYVKLKTLEKLLSQKAITGPDVTAHACNPSILGGLGRWITGGQEFKT
 SQANMVKPVSTKTTKILGMVVGVCNPSYLRG

A search of sequence databases reveals that the NOV28 amino acid sequence has 108/108 (100%) amino acids identical with ptrn:SPTREMBL-ACC:O60387 WUGSC:H_DJ0604G05.3 PROTEIN - Homo sapiens (Human). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV28 is expressed in at least Brain, Breast, Colon, Gall bladder, Germ Cell, Heart, Kidney, Liver, Ovary, Pancreas, Prostate, Stomach, Testis, Whole embryo, brain, breast, breast_normal, colon, colon_ins, head_neck, lung, nervous_tumor, prostate, prostate_normal, prostate_tumor, stomach. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV28 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 28C .

Table 28C. BLAST results for NOV28					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 3006230 gb AAC09485.1 (AC004522)	gap junction protein; similar to P36383 (PID:g544117) [Homo sapiens]	207	207/250 (82%)	207/250 (82%)	e-103
gi 17978264 ref NP_536698.1 (NM_080450)	gap junction membrane channel protein epsilon 1; connexin 29 [Mus musculus]	258	135/219 (61%)	167/219 (75%)	2e-75
gi 18566654 ref XP_095131.1 (XM_095131)	hypothetical protein XP_095131 [Homo sapiens]	222	100/115 (86%)	103/115 (88%)	3e-51
gi 6680007 ref NP_032148.1 (NM_008122)	gap junction membrane channel protein alpha 7; connexin 45 [Mus musculus]	396	96/250 (38%)	136/250 (54%)	5e-41
gi 4885169 ref NP_05488.1 (NM_005497)	gap junction protein, alpha 7, 45kD (connexin 45) [Homo sapiens]	396	96/250 (38%)	136/250 (54%)	8e-41

Table 28D lists the domain descriptions from DOMAIN analysis results against NOV28. This indicates that the NOV28 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 28D. Domain Analysis of NOV28	
gnl Pfam pfam00029 , connexin, Connexin.	
CD-Length = 218 residues, 90.4% aligned	
Score = 173 bits (438), Expect = 2e-44	

The disclosed NOV28 nucleic acid of the invention encoding a connexin-like protein includes the nucleic acid whose sequence is provided in Table 28A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 28A while still encoding a protein that maintains its connexin-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar

phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 16 percent of the bases may be so changed.

The disclosed NOV28 protein of the invention includes the connexin-like protein whose sequence is provided in Table 28B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 28B while still encoding a protein that maintains its connexin-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 0 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this connexin-like protein (NOV28) may function as a member of a "connexin family". Therefore, the NOV28 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV28 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the connexin-like protein (NOV28) may be useful in gene therapy, and the connexin-like protein (NOV28) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from Cardiomyopathy, Atherosclerosis, Hypertension, Congenital heart defects, Aortic stenosis, Atrial septal defect (ASD), Atrioventricular (A-V) canal defect, Ductus arteriosus, Pulmonary stenosis, Subaortic stenosis, Ventricular septal defect (VSD), valve diseases, Tuberosclerosis, Scleroderma, Obesity, Transplantation, Diabetes, Von Hippel-Lindau (VHL) syndrome, Pancreatitis, Obesity, Endometriosis, Fertility, Hemophilia, Hypercoagulation, Idiopathic thrombocytopenic purpura, Immunodeficiencies, Graft versus

host, Autoimmune disease, Renal artery stenosis, Interstitial nephritis, Glomerulonephritis, Polycystic kidney disease, Systemic lupus erythematosus, Renal tubular acidosis, IgA nephropathy, Hypercalceimia, Lesch-Nyhan syndrome, or other pathologies or conditions. The NOV28 nucleic acid encoding the connexin-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV28 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV28 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV28 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV29

A disclosed NOV29 nucleic acid of 1400 nucleotides (also referred to as CG57664-01) encoding a MHC Class I antigen-like protein is shown in Table 29A.

20

Table 29A. NOV29 nucleotide sequence (SEQ ID NO:67).

```
ATTCTCCCAACGCCAGGGATGGGGGTCATGGCTCCCCGAACCCTCCTCCTGCTGCTCTTGGGGGCCCT
GGCCCTGACCGAGACCTGGGCCGGTGAGTGCGGGGTCGGGAGGGAAAGGGCCTCTGCGGGGAGAAGCGAG
TGGCCCCGCCGGCCGGGGAGCCGCGCCTCAGCCTCTCCTCGCCTCCAGGCTCCCACTCCTTGAGGTATT
TCAGCACCGCAGTGTCCAGCCCCGCCGCGGGGAGCCCCGGTTTCATCGCCGTGGGTACGTGGACGACAC
AGAGTTGCGTGGGTTTCGACAGCGACTCCGTGAGTCCGAGGATGGAGCGGCGGGCGCCGTGGGTGGAGCAG
GAGGGGCTGGAGTATTGGGACCAGGAGACACGGAACGCCAAGGGCCACGCGCAGATTTACCGAGTGAACC
TGCGGACCTGCTCCGCTATTACAACCAGAGCGAGGCCGGTGGTTCTCACACCATCCAGAGGAAGCATGA
CTGCGACGTGGGCCCCGACAGGCGGGCCGACAGGCGCCTCCTCCGCAGGTATGAACAGTTTCGCTACGAT
GGCAAGGATTACATCGCCCTGAACGAGGACCTGCCCTCCTGGACCGCCGCGAACACAGCGGCTCAGATCT
CCCAGCACAAAGTGGGAAGCGGACAAATACTCAGAGCAGGTACGGGCTACCTGAGGGCAAGTGCATGGAG
TGGCGAGGGCAAGTGCATGGAGTGGCTCCGCAGACACCTGGAGAACGGGAAGGAGACGCTGCAGCGCGCG
TCAGATCCCCCAAAGGCACATGTGACCCAGCACCCCGTCTCTGACCATGAGGCCACCCTTGAGGTGCTGG
GCCCTGGGCCTCTACCCCTTGAGGTGCTGGGCCTTGGGCCTCTACCCCTGCGGAGATCACACTGACCTGGCA
GCAGGATGGGGAGGACAGACCCAGGACACGGAGCTTGTGGAGACCAGGCCTGCAGGGGACGGAACCTTC
CAGAAGTGGGTGGCTGTAGTGGTGCCTTCCGAGAGGAGCAGAGATACATGTGCCATGTGCAGCATGAGG
GGCTGCCAGAGCCCCCTCACCTGAGATGGCCCTCACCTCCCTCTCCTTTCCAGAGCCGCTCTCTCAGCC
CACCATCCCCATCGTGGGCATCGTTGCTGGCCTGTTTCTCCTTGGAGCTGTGGTCACTGGAGCTGTGGTT
GCTGCTGTGATGAAGAGGAAGAAAAGCTCAGGTAGGGAAGGGGTGAGAGGTGGGATCTGGGTTTCTTGT
TCCACTGTGGGTTTCAAGCCACAGGTAGAAATTGTGACTTGCTTCATCACTGGGAAGCACCGTCCACACAC
AGGCCGACCTAGCCTGGGGCCCTGTGTGCCAACACTTGCTCTTTTGTGAAGCACATGTGAAAACGAAGGA
```

In a search of public sequence databases, the NOV29 nucleic acid sequence, located on chromosome 6 has 332 of 353 bases (94%) identical to a gb:GENBANK-

ID:HUMHLA92|acc:M96338.1 mRNA from Homo sapiens (Homo sapiens HLA-92 gene sequence). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV29 polypeptide (SEQ ID NO:68) encoded by SEQ ID NO:67 has 452 amino acid residues and is presented in Table 29B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV29 has a signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.4600. The most likely cleavage site for a NOV29 peptide is between amino acids 24 and 25.

Table 29B. Encoded NOV29 protein sequence (SEQ ID NO:68).

```
MGVMAPRTL L L L L L L L L L L G A L A L T E T W A G E C G V G R E R A S A G R S E W P A R P G E P R L S L S S P P G S H S
L R Y F S T A V S Q P G R G E P R F I A V G Y V D D T E F V R F D S D S V S P R M E R R A P W V E Q E G L E Y W D Q E T
R N A K G H A Q I Y R V N L R T L L R Y Y N Q S E A G S S H T I Q R K H D C D V G P T G G P D R R L L R R Y E Q F A Y D
G K D Y I A L N E D L P S W T A A N T A A Q I S Q H K W E A D K Y S E Q V R A Y L R A S A W S G E G K C M E W L R R H L
E N G K E T L Q R A S D P P K A H V T Q H P V S D H E A T L E V L G P G P L P L R C W A L G L Y P A E I T L T W Q Q D G
E D Q T Q D T E L V E T R P A G D G T F Q K W A V V V P S G E E Q R Y M C H V Q H E G L P E P L T L R W P S P P S P F
P E P S S Q P T I P I V G I V A G L F L L G A V V T G A V A A V M K R K K S S G R E G V R G G I W V F L F H C G F Q A
T G R I V T C F I T G K H R P H T G R P S L G P C V P T L A L L
```

A search of sequence databases reveals that the NOV29 amino acid sequence has 158/223 (70%) identity and 177/223 (79%) similarity with SPTREMBL-ACC:Q9TPL2 MHC CLASS I ANTIGEN - Pan troglodytes (Chimpanzee). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV29 is expressed in at least Bone Marrow, Dermis, Hippocampus, Placenta, and Tonsils. Expression information was derived from the tissue sources of the sequences that were included in the derivation of the sequence of CG57664-01. The sequence is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:HUMHLA92|acc:M96338.1) a closely related Homo sapiens HLA-92 gene sequence homolog in species Homo sapiens :Lymphoblastoid cell line. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV29 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 29C.

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 8117819 gb AAF72785.1 AF168404.1 (AF168404)	MHC class I antigen [Pan troglodytes]	365	275/404 (68%)	300/404 (74%)	e-141
gi 8117799 gb AAF72776.1 AF168395.1 (AF168395)	MHC class I antigen [Pan troglodytes]	365	276/404 (68%)	301/404 (74%)	e-141
gi 2118771 pir I54493	MHC class I histocompatibilit y antigen HLA-A alpha chain precursor human	365	271/404 (67%)	299/404 (73%)	e-141
gi 8117808 gb AAF72780.1 AF168399.1 (AF168399)	MHC class I antigen [Pan troglodytes]	365	275/404 (68%)	300/404 (74%)	e-140
gi 2251146 emb CAA65501.1 (X96724)	human leukocyte antigen [Homo sapiens]	365	270/404 (66%)	300/404 (73%)	e-140

Tables 29D-E list the domain descriptions from DOMAIN analysis results against NOV29. This indicates that the NOV29 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 29D. Domain Analysis of NOV29

[gnl|Pfam|pfam00129](#), MHC_I, Class I Histocompatibility antigen, domains alpha 1 and 2

CD-Length = 179 residues, 96.6% aligned

Score = 248 bits (634), Expect = 4e-67

Table 29E. Domain Analysis of NOV29

[gnl|Smart|smart00407](#), IGc1, Immunoglobulin C-Type

CD-Length = 75 residues, 96.0% aligned

Score = 59.7 bits (143), Expect = 4e-10

10 The major histocompatibility complex (MHC) encodes the class I and class II families
of glycoproteins that present peptides for immunorecognition by cytotoxic and helper T
lymphocytes, respectively. Class I molecules bind peptides generated by degradation of
proteins intracellularly, whereas class II molecules associate mainly with peptides derived
from endocytosed extracellular proteins. Two genes encode components of the proteasome
complex, which degrades cytosolic proteins and may generate antigenic peptides. Two closely
15 linked genes, PSF1 and PSF2, encode subunits of a transporter, which presumably translocates

peptides into an exocytic compartment where they associate with class I molecules. The location of these genes in the MHC in close linkage to the class I and class II gene families suggests that they coevolved to optimize functional interactions.

The disclosed NOV29 nucleic acid of the invention encoding a MHC Class I antigen-like protein includes the nucleic acid whose sequence is provided in Table 29A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 29A while still encoding a protein that maintains its MHC Class I antigen-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 6 percent of the bases may be so changed.

The disclosed NOV29 protein of the invention includes the MHC Class I antigen-like protein whose sequence is provided in Table 29B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 29B while still encoding a protein that maintains its MHC Class I antigen-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 30 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this MHC Class I antigen-like protein (NOV29) may function as a member of a "MHC Class I antigen family". Therefore, the NOV29 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene

therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV29 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the MHC Class I antigen-like protein (NOV29) may be useful in gene therapy, and the MHC Class I antigen-like protein (NOV29) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, Tonsillitis, Hemophilia, hypercoagulation, Idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, Graft versus host, or other pathologies or conditions. The NOV29 nucleic acid encoding the MHC Class I antigen-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV29 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV29 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV29 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV30

A disclosed NOV30 nucleic acid of 1225 nucleotides (also referred to as CG57666-01) encoding a MHC Class I antigen-like protein is shown in Table 30A. The start and stop codons are in bold letters.

Table 30A. NOV30 nucleotide sequence (SEQ ID NO:69).

ACGCCGAGGATGGGGTCATGGCGTCCCAAACCTCCTCCTGCTGCTCTTGGGGGCCCTGGCCCTGACCGA
GACCTGGGCGGGTACCCACTCCATAAGGTATTTTCAGCACCGCCGTGTCCCGGCCGGGTTCGCGGGGAGCCC
CGGGGTACCCACTCCATAAGGTATTTTCAGCACCGCCGTGTCCCGGCCGGGTTCGCGGGGAGCCCCGGGTACA
TCGCAGTGGGCTACGTGGACGACACGAGTTCGTGCGGTTTCGACAGCGACGCGGCGACTCCGAGGATGGA
GCCGACAGGCGCCGTGGTTGGAGCAGGAGGGACCGGAGTATTGGGACCGGAGCACACCGAACATCAGGCCC
GCGCACAGACTGACAAGAGTGAACCTGCCCATGCCGCGCGCTACTACCACAGAGCGGGTCTAACACCC
TCCAGATAATGTATGGCTGCGACTTGGGGCTGGAAGGGCGCCTCCTCCGCGGGTATGAACAGCAGCCAA
CGATGGCAAAGATTACATCGCCCTAAACGAGGACCTGAGCTCTTGGACCGCGCGGCCATGGCGGCTCAG
ATTACCCAGCGCAAGTGGGAGGCGGCCCATGAGGCGGAGCAGCAGAGAGCCTACCTGGAGGGCAGCTGCG
TGGAGTGGCTCCGCAGATACCTGGAGAACGGGAAGGAGACGCTGCAGCGCACTACCCCCCCCCCAAGAC
ACATATGATCCACCATTCCGTCTCTGACTATAAGGCCACCCTGAGATGCTGGGCCCTGGGCTTCTACCCT
GTGGAGATCACACTGACCTGGCAGCAGGATGGAGAGGACCAGACTCAGGACATGGAGCTTGTAGAGACCA
GGCCTGCAGGGGATGGAACCTCCAGAAGTGGGCGAGCTGTGGTGGTGCCTTCTGGAGAGGAACAGAGATA
CATGTGCCATGTGCAGCATGAGGGGTGCCCAAGCCCTCACCTGAGATGGGAGCAGTCTTCTCAGCCC
ACCATCCCCATCGTGGGTATCGTTGCTGGCCTGGTTCTCCTTGGAGCTGTAGTCACTGGAGCTGTGGTTT
CTGCTGTGATGTGCAGGAAGAACTCATTTTGTCTACCCAGGCAGCAACCATGCGCAGGGTCTGATGT
GTCTCTCAGGCTTGTAAAGGTGAGACGCTGGGGGACCTGATGTGTGGGGGTGTTGGGGGCAATAGTGG
ATGCAGCTGTGCTATGGGGTTTCTTTGAATTGGAT

In a search of public sequence databases, the NOV30 nucleic acid sequence, located on chromosome 6 has 265 of 271 bases (97%) identical to a gb:GENBANK-
ID:AF055066|acc:AF055066.1 mRNA from Homo sapiens (Homo sapiens MHC class 1
5 region). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV30 polypeptide (SEQ ID NO:70) encoded by SEQ ID NO:69 has 389 amino acid residues and is presented in Table 30B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV30 has a signal peptide and is
10 likely to be localized at the plasma membrane with a certainty of 0.4600. The most likely cleavage site for a NOV30 peptide is between amino acids 21 and 22.

Table 30B. Encoded NOV30 protein sequence (SEQ ID NO:70).

MASQTL L L L L L L L L L L G A L A L T E T W A G T H S I R Y F S T A V S R P G R G E P R G T H S I R Y F S T A V S R P G R G
E P R Y I A V G Y V D D T Q F V R F D S A A T P R M E P Q A P W L E Q E G P E Y W D R S T P N I R P A H R L T R V N L
P M P R R Y Y H Q S G S N T L Q I M Y G C D L G L E G R L L R G Y E Q H A N D G K D Y I A L N E D L S S W T A A M A A
Q I T Q R K W E A A H E A E Q Q R A Y L E G T C V E W L R R Y L E N G K E T L Q R T T P P P K T H M I H S V S D Y K A
T L R C W A L G F Y P V E I T L T W Q Q D G E D Q T Q D M E L V E T R P A G D G N F Q K W A A V V V P S G E E Q R Y M C
H V Q H E G L P K P L T L R W E Q S S Q P T I P I V G I V A G L V L L G A V V T G A V V S A V M C R K N S F C S T P G S
N H A Q G S D V S L T A C K G E T L G D L M C G G C W G Q

A search of sequence databases reveals that the NOV30 amino acid sequence has 258/338 (76%) identity and 284/338 (84%) similarity with SPTREMBL-ACC:Q31602 MHC
15 CLASS I ANTIGEN - Homo sapiens. Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV30 is expressed in at least Bone Marrow, Dermis, Hippocampus, Placenta, Tonsils. This information was derived by determining the tissue sources of the sequences that

were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV30 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 30C.

5

Table 30C. BLAST results for NOV30					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
<u>gi 8571968 gb AAF76942.1 AF179640.1</u> (AF179640)	MHC class I antigen [Pan troglodytes]	365	283/383 (73%)	313/383 (80%)	e-155
<u>gi 8571970 gb AAF76943.1 AF179641.1</u> (AF179641)	MHC class I antigen [Pan troglodytes]	363	282/383 (73%)	312/383 (80%)	e-155
<u>gi 8117808 gb AAF72780.1 AF168399.1</u> (AF168399)	MHC class I antigen [Pan troglodytes]	365	282/383 (73%)	312/383 (80%)	e-155
<u>gi 2118771 pir I54493</u>	MHC class I histocompatibilit y antigen HLA-A alpha chain precursor- human	365	280/383 (73%)	312/383 (80%)	e-154
<u>gi 6049049 gb AAF02442.1 </u> (AF115463)	MHC class I antigen [Pan troglodytes]	365	281/383 (73%)	311/383 (80%)	e-153

Tables 30D-E list the domain descriptions from DOMAIN analysis results against NOV30. This indicates that the NOV30 sequence has properties similar to those of other proteins known to contain this domain.

10

Table 30D. Domain Analysis of NOV30
<u>gnl Pfam pfam00129</u> , MHC_I, Class I Histocompatibility antigen, domains alpha 1 and 2.
CD-Length = 179 residues, 100.0% aligned
Score = 245 bits (626), Expect = 3e-66

Table 30E. Domain Analysis of NOV30
<u>gnl Smart smart00407</u> , IGc1, Immunoglobulin C-Type
CD-Length = 75 residues, 100.0% aligned
Score = 75.5 bits (184), Expect = 5e-15

The disclosed NOV30 nucleic acid of the invention encoding a MHC Class I antigen-like protein includes the nucleic acid whose sequence is provided in Table 30A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may

15



be changed from the corresponding base shown in Table 30A while still encoding a protein that maintains its MHC Class I antigen-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 3 percent of the bases may be so changed.

The disclosed NOV30 protein of the invention includes the MHC Class I antigen-like protein whose sequence is provided in Table 30B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 30B while still encoding a protein that maintains its MHC Class I antigen-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 24 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this MHC Class I antigen-like protein (NOV30) may function as a member of a “MHC Class I antigen family”. Therefore, the NOV30 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV30 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the MHC Class I antigen-like protein (NOV30) may be useful in gene therapy, and the MHC Class I antigen-like protein (NOV30) may be useful when administered to a subject in need thereof. By way of

nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, Tonsillitis, Hemophilia, hypercoagulation, Idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, Graft versus host, or other pathologies or conditions. The NOV30 nucleic acid encoding the MHC Class I antigen-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV30 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV30 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV30 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV31

A disclosed NOV31 nucleic acid of 1159 nucleotides (also referred to as CG57668-01) encoding a MHC Class I antigen-like protein is shown in Table 31A. The start and stop codons are in bold letters.

Table 31A. NOV31 nucleotide sequence (SEQ ID NO:71).

TCTCCCCAGACGCCGAGGATGGTGCTCATGGCGCCCCGAACCCTCCTCCTGCTGCTCTCAGGGGGCCCTGA
 CCCAGACCTGGGCGCGTTCCCACTCCATGAGGTATTTCTACACCACCATGTCCCGGCCCGGCCGCGGGGA
 GCCCCGCTTCATCTCCGTCGGCTACGTGGACTATACGCAGTTCGTGCGGTTTCGACAGCGACGACGCGAGT
 CCGAGAGAGGAGCCGCGGGCGCGTGGATGGAGCGGGAGGGGCCGAGTATTGGGACCGGAACACACAGA
 TCTGCAAGGCCCCAAGCACGGACTGAACGAGAGAACCTGCGGATCGCGCTCCGCTACTACAACAGAGCGGA
 GGGCGGTGGTTCCCAACCATGCAGGTGATGTATGGCTGCGACGTGGGGCCCCGACGGGCGCTTCCTCCGC
 GGGTATGAACAGCACGCCTACGACGGCAAGGATTACATCGCTCTGAACGAGGACCTGCGCTCCTGGACCG
 CGGCGGACATGGCAGCTCAGATACCAAGCGCAAGTGGGAGGCGGCCCGTGTGGCGGAGCAGCTGAGAGC
 CTACCTGGAGGGCGAGTTCGTGGAGTGGCTCCGCAGATACCTGGAGAACGGGAAGGAGACGCTGCAGCGC
 CGCTCAGACCCCCCAAGACACATATGACCCACTACCCCATCTCTGACCATGAGGCCACCCTGAGGTGCT

Table 31C. BLAST results for NOV31					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 232260 sp P01893 HLAH HUMAN	HLA class I histocompatibility antigen, alpha chain H precursor (HLA-AR) (HLA-12.4)	362	330/376 (87%)	339/376 (89%)	e-167
gi 70075 pir HLHU12	MHC class I histocompatibility antigen HLA alpha chain precursor (clone pHLA 12.4) - human	359	329/373 (88%)	337/373 (90%),	e-166
gi 2118771 pir I54493	MHC class I histocompatibility antigen HLA-A alpha chain precursor - human	365	303/376 (80%)	326/376 (86%),	e-164
gi 4164602 gb AAD05568.1 (AF116214)	MHC class I antigen heavy chain [Homo sapiens]	365	302/376 (80%)	325/376 (86%),	e-163
gi 915219 gb AAA73518.1 (U25971)	MHC class I antigen HLA-A2407 [Homo sapiens]	365	302/376 (80%)	326/376 (86%),	e-163

Tables 31D-E list the domain descriptions from DOMAIN analysis results against NOV31. This indicates that the NOV31 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 31D. Domain Analysis of NOV31
gnl Pfam pfam00129 , MHC_I, Class I Histocompatibility antigen, domains alpha 1 and 2 CD-Length = 179 residues, 99.4% aligned Score = 284 bits (727), Expect = 5e-78

Table 31E. Domain Analysis of NOV31
gnl Smart smart00407 , IGc1, Immunoglobulin C-Type CD-Length = 75 residues, 98.7% aligned Score = 77.8 bits (190), Expect = 1e-15

The disclosed NOV31 nucleic acid of the invention encoding a MHC Class I antigen-like protein includes the nucleic acid whose sequence is provided in Table 31A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may

10

be changed from the corresponding base shown in Table 31A while still encoding a protein that maintains its MHC Class I antigen-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 22 percent of the bases may be so changed.

The disclosed NOV31 protein of the invention includes the MHC Class I antigen-like protein whose sequence is provided in Table 31B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 31B while still encoding a protein that maintains its MHC Class I antigen-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 9 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this MHC Class I antigen-like protein (NOV31) may function as a member of a "MHC Class I antigen family". Therefore, the NOV31 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV31 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the MHC Class I antigen-like protein (NOV31) may be useful in gene therapy, and the MHC Class I antigen-like protein (NOV31) may be useful when administered to a subject in need thereof. By way of

nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, Tonsillitis, Hemophilia, hypercoagulation, Idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, Graft versus host as well, or other pathologies or conditions. The NOV nucleic acid encoding the MHC Class I antigen-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV31 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV31 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV31 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV32

A disclosed NOV32 nucleic acid of nucleotides (also referred to as CG57660-01) encoding a retinoic acid receptor responder-like protein is shown in Table 32A. The start and stop codons are in bold letters.

Table 32A. NOV32 nucleotide sequence (SEQ ID NO:73).

```
AGGGGAAGCATGAGACGGCTGCGGATCTCGCTGGCCCCGTGGGTGGGCGCGGGGACGCGGGAGGGGCCG
AGCTCACGGGGCCAGCGCCGGGGCCTGCAGGTGGCCCTGGAGGAATCTGCAAGCACCCGCCCCGTGCAGCG
GGCCTTCCGGGAGACCAGTGTGGACAGCGCCCTGGACACGCCCTTCCCAGCTGGAACATCTGTGAGGCTG
GAATTTAAGCTCCGGCAGACAAGCGGCTGGAGGAAGGCTGGAAGAAACCAAGTGCAAAGCCAGCCCG
AGAGGAGGAAACAGAAATGCCTGACCTGCGTCAAATGGACTGTGAGGATAAGGTTCTGGGCAGGATGGT
TCGCTGCCCTCCAGAGACGCAGACTCGGCGGGAGCCTGAGGAGCACCAGGGGGCCGGGTGCAGCCCGGCG
GAGCGGGCGGGGAGGACCCACGGCGGAGCGGGCGGGGAGGACCCACGGCTGCCGCTTCCCTGCACGGT
TCGCCTCCTCCAAGGCCCGGCCCCAGCGGAGCCCTAGCGCTGAATCGCATGGCGCCCCCTGGAGCCCTG
GCGGG
```

In a search of public sequence databases, the NOV32 nucleic acid sequence, located on chromosome 4 has 443 of 451 bases (98%) identical to a gb:GENBANK-ID:AF146191|acc:AF146191.1 mRNA from Homo sapiens (Homo sapiens FRG1 (FRG1) gene, complete cds; 5S ribosomal RNA gene, complete sequence; TUB4q and TIG2 pseudogenes, complete sequence). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV32 polypeptide (SEQ ID NO:74) encoded by SEQ ID NO:73 has amino acid residues and is presented in Table 32B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV32 has a signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.8800.

Table 32B. Encoded NOV32 protein sequence (SEQ ID NO:74).

MRRLRISLAPWVGAGDAGGAELTGPAAGPAGGPGGICKHPPVQRAFRETSDSALDTPFP
AGTSVRLEFKLRQTSGWRKAWKKPKCKAQPERRKQKCLTCVKMDCEDKVLGRMVRCPPET
QTRREPEEHQAGCSPAERAGRTPRRSGRGGPHGCRFPARFASSKARPPAEP

A search of sequence databases reveals that the NOV32 amino acid sequence has 94/168 (55%) identity and 109/168 (64%) similarity with ptnr:SWISSPROT-ACC:Q99969 RETINOIC ACID RECEPTOR RESPONDER PROTEIN 2 PRECURSOR (TAZAROTENE-INDUCED GENE 2 PROTEIN) (RAR-RESPONSIVE PROTEIN TIG2)- Homo sapiens. Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV32 is expressed in at least Adipose, Adrenal gland, Breast, Colon, Esophagus, Eye, Heart, Kidney, Liver, Lung, Ovary, Parathyroid, Placenta, Prostate, Stomach, Testis, Uterus, Whole embryo, bladder tumor, brain, cervix, colon, head and neck, kidney, lung, muscle and ovary. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV32 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 32C.

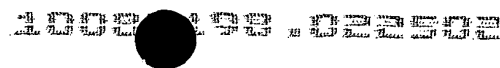
Table 32C. BLAST results for NOV32

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 4506427 ref NP_02880.1 (NM_002889)	retinoic acid receptor responder (tazarotene induced) 2 [Homo sapiens]	163	94/169 (55%)	109/169 (63%)	8e-37

<u>gi 12832179 dbj BAB21997.1</u> (AK002298)	homolog to RETINOIC ACID RECEPTOR RESPONDER PROTEIN 2 PRECURSOR (TAZAROTENE-INDUCED GENE 2 PROTEIN) (RAR-RESPONSIVE PROTEIN TIG2)~putative [Mus musculus]	162	72/169 (42%)	94/169 (55%) ,	3e-22
<u>gi 17436162 ref XP_067978.1</u> (XM_067978)	similar to retinoic acid receptor responder (tazarotene induced) 2 [Homo sapiens]	206	73/74 (98%)	74/74 (99%)	1e-20
<u>gi 18585954 ref XP_091375.1</u> (XM_091375)	hypothetical protein XP_091375 [Homo sapiens]	322	52/79 (65%)	59/79 (73%)	2e-18
<u>gi 17488051 ref XP_064055.1</u> (XM_064055)	similar to Fibroblast growth factor receptor 3 precursor (FGFR-3) (Heparin-binding growth factor receptor) [Homo sapiens]	296	28/54 (51%)	32/54 (58%) ,	2e-5

Retinoids exert their biologic effects through two families of nuclear receptors, retinoic acid receptors (RARs) and retinoid X receptors (RXRs), which belong to the superfamily of steroid/thyroid hormone nuclear receptors. The retinoid-mediated up-regulation in the expression of TIG2 was confirmed by Northern blot analysis. Upon sequencing, TIG2 was found to be a cDNA whose complete sequence was not in the GenBank and EMBL data bases. The TIG2 cDNA is 830 bp long and encodes a putative protein product of 164 amino acids. TIG2 is neither expressed nor induced by tazarotene in primary keratinocyte and fibroblast cultures. Thus, TIG2 is expressed and induced by tazarotene only when keratinocytes and fibroblasts form a tissue-like 3-dimensional structure. RAR-specific retinoids increase TIG2 mRNA levels. In contrast, neither RXR-specific retinoids nor 1,25-dihydroxyvitamin D3 increased TIG2 levels. TIG2 is expressed at high levels in nonlesional psoriatic skin but at lower levels in the psoriatic lesion and that its expression is up-regulated in psoriatic lesions after topical application of tazarotene.

The disclosed NOV32 nucleic acid of the invention encoding a retinoic acid receptor responder-like protein includes the nucleic acid whose sequence is provided in Table 32A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 32A while still encoding a protein that maintains its retinoic acid receptor responder-like activities and physiological



functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures
5 include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic
10 acids, and their complements, up to about 2 percent of the bases may be so changed.

The disclosed NOV32 protein of the invention includes the retinoic acid receptor responder-like protein whose sequence is provided in Table 32B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 32B while still encoding a protein that maintains its retinoic acid
15 receptor responder-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 45 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this retinoic acid
20 receptor responder-like protein (NOV32) may function as a member of a “retinoic acid receptor responder family”. Therefore, the NOV32 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target,
25 antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV32 nucleic acids and proteins of the invention are useful in potential
30 therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the retinoic acid receptor responder-like protein (NOV32) may be useful in gene therapy, and the retinoic acid receptor responder-like protein (NOV32) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy

for treatment of patients suffering from Cardiomyopathy, Atherosclerosis, Hypertension, Congenital heart defects, Aortic stenosis, Atrial septal defect (ASD), Atrioventricular (A-V) canal defect, Ductus arteriosus, Pulmonary stenosis, Subaortic stenosis, Ventricular septal defect (VSD), valve diseases, Tuberous sclerosis, Scleroderma, Obesity, Transplantation, Diabetes, Von Hippel-Lindau (VHL) syndrome, Pancreatitis, Obesity, Endometriosis, Fertility, Hemophilia, Hypercoagulation, Idiopathic thrombocytopenic purpura, Immunodeficiencies, Graft versus host, Autoimmune disease, Renal artery stenosis, Interstitial nephritis, Glomerulonephritis, Polycystic kidney disease, Systemic lupus erythematosus, Renal tubular acidosis, IgA nephropathy, Hypercalcaemia, Lesch-Nyhan syndrome, or other pathologies or conditions. The NOV32 nucleic acid encoding the retinoic acid receptor responder-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV32 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV32 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV32 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV33

A disclosed NOV33 nucleic acid of 1706 nucleotides (also referred to as) encoding a PHOSPHATIDYLINOSITOL 4-PHOSPHATE 5-KINASE -like protein is shown in Table 33A. The start and stop codons are in bold letters.

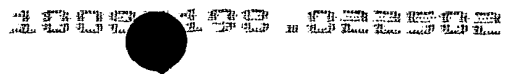
Table 33A. NOV33 nucleotide sequence (SEQ ID NO:75).

<p>CTGCCAAGATGGCGTCGGCCTCCTCCCAACCGTCGTTGGCGGTCGGTTTTTCATCCTTTGATCCCGGGC CCCTTCCTGTACCGCGTCCTCAGCATCTGGAATCTTGAGCCCCACGGCATCTGAGGTGCCTTATGCCTCT GGCATGCCCATCAAGAAAACAGGCCATCGAGGTGTCGATTTCCTCAGGAGAGACAACATATAAAAAGACAA CCTCAACAGCCTTGAAAGGTGCCATCCAGTTAGGCATTACTTACACTGTGGGGAGCCTGAGTACCAACC AGAGCGTGATGTCCTCATGCAAGATTTCTACGTGGTGGAGAGTATCTTCTCCCCAGTGAAGGGAGCAAC CTGACCCCTGCTCATCACTACAATGCCTTCCGTTTCAAGACCTATGCGCCGGTTGCCTTCCGCTACTTTC GGGAGCTATTTGGTATCCCGCCGATGATTACTTGTGCTCCCTCTGCAGTGAGCCGCTGATTGAACTCTG</p>
--

NOV33 is expressed in at least Adrenal gland, Aorta, B-cells, Blood, Bone, Brain, Breast, CNS, Colon, Ear, Esophagus, Eye, Gall bladder, Germ Cell, Head and neck, Heart, Kidney, Larynx, Liver, Lung, Lymph, Marrow, Muscle, Neural, Omentum, Ovary, Pancreas, Parathyroid, Peripheral nervous system, Placenta, Pooled, Prostate, Skin, Small intestine, Spleen, Stomach, Synovial membrane, Testis, Tissue culture, Tonsil, Uterus, Whole embryo, and adrenal gland. Expression information was derived from the tissue sources of the sequences that were included in the derivation of the sequence of CG57672-01. The sequence is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:HSU78575|acc:U78575.1) a closely related Human 68 kDa type I phosphatidylinositol-4-phosphate 5-kinase alpha mRNA, clone PIP5KIa1, complete cds homolog in species Homo sapiens :fetal brain. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV33 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 33C.

Table 33C. BLAST results for NOV33					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 4505815 ref NP_03548.1 (NM_003557)	phosphatidylinositol-4-phosphate 5-kinase, type I, alpha [Homo sapiens]	549	478/552 (86%)	496/552 (89%)	0.0
gi 1743873 gb AAC50911.1 (U78576)	68 kDa type I phosphatidylinositol-4-phosphate 5-kinase alpha [Homo sapiens]	562	478/565 (84%)	496/565 (87%)	0.0
gi 1743875 gb AAC50912.1 (U78577)	68 kDa type I phosphatidylinositol-4-phosphate 5-kinase alpha [Homo sapiens]	500	439/551 (79%)	455/551 (81%)	0.0
gi 6679331 ref NP_032873.1 (NM_008847)	phosphatidylinositol-4-phosphate 5-kinase, type 1 beta; PI4P5K-I[b] [Mus musculus]	546	434/552 (78%)	470/552 (84%)	0.0
gi 14745097 ref XP_018166.2 (XM_018166)	similar to phosphatidylinositol-4-phosphate 5-kinase, type I, alpha (H. sapiens) [Homo sapiens]	435	361/408 (88%)	381/408 (92%)	0.0



databases with peptide sequences obtained from the 68-kD type I PIP5K purified from bovine erythrocytes, Loijens and Anderson (1996) identified a human EST encoding PIP5K1A, which they called PIP5KI-alpha. They screened a human fetal brain cDNA library and isolated full-length PIP5K1A cDNAs. The deduced 549-amino acid protein has the conserved kinase
5 homology domain of PIP5K family members. Within this domain, PIP5K1A shows 83% and 35% amino acid identity with PIP5K1B and PIP5K2A, respectively. Overall, the PIP5K1A and PIP5K1B proteins are 64% identical. Recombinant PIP5K1A expressed in bacteria had a molecular mass of approximately 66.3 kD by Western blot analysis. The authors isolated additional PIP5K1A cDNAs which they suggested represent splicing isoforms. Northern blot
10 analysis detected a major 4.2-kb PIP5K1A transcript which had a wide tissue distribution.

The disclosed NOV33 nucleic acid of the invention encoding a PHOSPHATIDYLINOSITOL 4-PHOSPHATE 5-KINASE-like protein includes the nucleic acid whose sequence is provided in Table 33A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the
15 corresponding base shown in Table 33A while still encoding a protein that maintains its PHOSPHATIDYLINOSITOL 4-PHOSPHATE 5-KINASE-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally
20 includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense
25 binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 8 percent of the bases may be so changed.

The disclosed NOV33 protein of the invention includes the PHOSPHATIDYLINOSITOL 4-PHOSPHATE 5-KINASE-like protein whose sequence is provided in Table 33B. The invention also includes a mutant or variant protein any of whose
30 residues may be changed from the corresponding residue shown in Table 33B while still encoding a protein that maintains its PHOSPHATIDYLINOSITOL 4-PHOSPHATE 5-KINASE-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 14 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this PHOSPHATIDYLINOSITOL 4-PHOSPHATE 5-KINASE-like protein (NOV33) may
 5 function as a member of a "PHOSPHATIDYLINOSITOL 4-PHOSPHATE 5-KINASE family". Therefore, the NOV33 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target
 10 (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV33 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the
 15 PHOSPHATIDYLINOSITOL 4-PHOSPHATE 5-KINASE-like protein (NOV33) may be useful in gene therapy, and the PHOSPHATIDYLINOSITOL 4-PHOSPHATE 5-KINASE-like protein (NOV33) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for
 20 treatment of patients suffering from Cardiomyopathy, Atherosclerosis, Hypertension, Congenital heart defects, Aortic stenosis, Atrial septal defect (ASD), Atrioventricular (A-V) canal defect, Ductus arteriosus, Pulmonary stenosis, Subaortic stenosis, Ventricular septal defect (VSD), valve diseases, Tuberous sclerosis, Scleroderma, Obesity, Transplantation, Adrenoleukodystrophy, Congenital Adrenal Hyperplasia, Hemophilia,
 25 Hypercoagulation, Idiopathic thrombocytopenic purpura, Immunodeficiencies, Graft versus host, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, or other pathologies or conditions. The
 30 NOV33 nucleic acid encoding the PHOSPHATIDYLINOSITOL 4-PHOSPHATE 5-KINASE-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV33 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV33 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV33 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV34

A disclosed NOV34 nucleic acid of 1316 nucleotides (also referred to as CG57680-01) encoding a Cyclophilin-type peptidyl-prolyl cis-trans isomerase-like protein is shown in Table 34A. The start and stop codons are in bold letters.

Table 34A. NOV34 nucleotide sequence (SEQ ID NO:77).

TGGGTGCCATGGCGGTTCTACTGGAGACTACTGTGGCAATGTGGTTGTCAATTTGCACACTGAGCAGCA
GCCTTGCAACTGTGAACCTTTTTGAGAGCAGGTACCACAGTTTAATGGCATTAAATTTCTTGAGATATTAC
AAAATAAAATATTACAGTTATTGCCCTTATTACAGTATACAAAGGTATTTTATCATACAACTGTTGATC
CTACAGGGACTGGTTCATGGAGGAGAGTCTATTTTGGCCTAGGATTGTATGGTGTATCAAGCAAGCTTTT
TGAGACAGAAAACGTCCCAAGAATTAAGCACAAGAAGAAGGGCACAATGTCCATGGTGAATAATGACAGT
GATCAACATGGATCTCAGTTTCTTATCACTACAGGAGAAAATCTAGATTACCTTGATGGTACCCATACAG
TATTTGGTGAGGTGACAGAAGGCATTGACATAATTAAGAAAATAAATGAGACCTTTGTTGACAAGGACTT
TGTACCATATCAGGATATCAGGATAAATTATATAGTGAATTTAGATGGTCCATTTGATGACATTCCTGAT
TTATTAATCCCTGATCAATCACCAGAACCTACAAGGGAACAATTAAGAGTGGTAGAGTTGACACAAATG
AAGAAATGATCATTTCAAACGAAGGTCAGCCGAAGAAGTAGAAGAAATAAAGGCAGAAAAAGAAGCTAA
AACTCAGGCTTTACTTTTAGAGATGGTGGGAGACCTACCTGATGCAGATATTAACCTCCGGAAAAATCT
GTGTTTGTATGCAAATTGAATCCAGTGACCACAGATGAGGATCTGGATATAATACTCTCTAGATTTGGGC
CAATAAGAAGTTGTGAAGTTATCTGGGACTGGAAGACAGGAGAAATCCTCTGTTATTTCTTTCTTTCTT
CTATGCTTTTATTGAATTTGAAAAGGAAGAAGATTATGAGAAAGCCTTCTTCAAAATGGACAATATACTT
ATAGATGACAGAAGAAAACATGGATTTGCCAGTCTGTTACAAAGGTTAAATGGAAGGAAAAAGTGGGAAA
TACACCAACAGCCGGGGCGCAGCCACGCCGCCGCCGCCGCCGCCGCCGCTCCCGCTCCCGCGGGCGGCG
ACGGCGGGCGGGGACCCCGCGGCGCTGCGCCTCATCCTCTGCGACGGGGGATGAGGGGTCTGAGAGGAAC
TGGAGGAGGAGGAGGAGGAGGCAGTGGACATGGTGGCTCTGCGGGCCCTGGACGCCGCCGCCGACACGG
AGGCGCGGGGCGGAGGCGCGAGCGTGTGGGAGCCGGCGGATAGTGCGTGCGA

In a search of public sequence databases, the NOV34 nucleic acid sequence, located on chromosome 7 has 149 of 235 bases (63%) identical to a gb:GENBANK-ID:A45258|acc:A45258.1 mRNA from human herpesvirus 2 (Sequence 2 from Patent WO9516779). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV34 polypeptide (SEQ ID NO:78) encoded by SEQ ID NO:77 has 432 amino acid residues and is presented in Table 34B using the one-letter amino acid code.

Signal P, Psort and/or Hydropathy results predict that NOV34 has a signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.9820.

Table 34B. Encoded NOV34 protein sequence (SEQ ID NO:78).

MAVLLETTVGNNVNLHTEQQPCNCELFSRYHSLMAFNFLRYKIKYYSYCLIHSIQRY
FIIQTVDPTGTGHGGESIFGLGLYGDQASFFETENVPRIKHKKGTMSMVNNDSDQHGSQ
FLITTGENLDYLDGTHTVFGEVTEGIDI IKKINETFVDKDFVPYQDIRINYIVILDGPF
DIPDLLIPDQSPEPTREQLKSGRVDTNEEIDHFKRRSAEEVEEIKAEKEAKTQALLLEMV
GDLPDADIKPPEKSVFVCKLNPVTTDEDLDIILSRFGPIRSCEVIWDWKTGEILCYFFLS
FYAFIEFEKEEDYEKAFKMDNILIDRRKHGFASLLQRLNGRKKWEIHQQPGRSPRRRR
RPHRSRSPRRRRRAGTPRRCASSSATGDEGSERNWRRRRRRQWTWWLCGPWTPPADTEAR
GGRRASVWEFAR

A search of sequence databases reveals that the NOV34 amino acid sequence has 263/345 (76%) identity and 291/345 (84%) similarity with TREMBLNEW-ACC: BAB30711 6 DAYS NEONATE HEAD CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY, CLONE:5430431E21, FULL INSERT SEQUENCE – Mus musculus. Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV34 is expressed in at least Adrenal Gland/Suprarenal gland, Bone, Bone Marrow, Brain, Colon, Hair Follicles, Heart, Hippocampus, Kidney, Liver, Lung, Lymphoid tissue, Pancreas, Peripheral Blood, Prostate, Salivary Glands, Small Intestine, Testis, Tonsils, Uterus. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV34 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 34C.

Table 34C. BLAST results for NOV34

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 12849069 dbj BAB28194.1 (AK012371)	homolog to DJ12G14.1 (NOVEL CYCLOPHILIN TYPE PEPTIDYL-PROLYL CIS-TRANS ISOMERASE) (FRAGMENT) ~putative [Mus musculus]	382	261/331 (78%)	284/331 (84%)	e-138
gi 12847571 dbj BAB27623.1 (AK011443)	homolog to DJ12G14.1 (NOVEL CYCLOPHILIN TYPE PEPTIDYL-PROLYL CIS-TRANS ISOMERASE) (FRAGMENT) ~putative [Mus musculus]	418	261/331 (78%)	284/331 (84%)	e-137

<u>gi 12856573 dbj BAB</u> <u>30711.1</u> (AK017370)	homolog to DJ12G14.1 (NOVEL CYCLOPHILIN TYPE PEPTIDYL-PROLYL CIS-TRANS ISOMERASE) (FRAGMENT)-putati ve [Mus musculus]	460	261/331 (78%)	284/331 (84%)	e-136
<u>gi 12852237 dbj BAB</u> <u>29330.1</u> (AK014406)	homolog to DJ12G14.1 (NOVEL CYCLOPHILIN TYPE PEPTIDYL-PROLYL CIS-TRANS ISOMERASE) (FRAGMENT)-putati ve [Mus musculus]	492	259/331 (78%)	284/331 (85%)	e-135
<u>gi 18088111 gb AAH2</u> <u>0986.1 AAH20986</u> (BC020986)	protein for MGC:9727) [Homo sapiens]	492	263/331 (79%)	284/331 (85%)	e-135

Tables 34D-E list the domain descriptions from DOMAIN analysis results against NOV34. This indicates that the NOV34 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 34D. Domain Analysis of NOV34

gnl|Pfam|pfam00160, pro_isomerase, Cyclophilin type peptidyl-prolyl
cis-trans isomerase

CD-Length = 162 residues, 88.9% aligned

Score = 91.7 bits (226), Expect = 8e-20

Table 34E. Domain Analysis of NOV34

[gnl|Pfam|pfam00076](#), rrm, RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain). The RRM motif is probably diagnostic of an RNA binding protein. RRMs are found in a variety of RNA binding proteins, including various hnRNP proteins, proteins implicated in regulation of alternative splicing, and protein components of snRNPs. The motif also appears in a few single stranded DNA binding proteins. The RRM structure consists of four strands and two helices arranged in an alpha/beta sandwich, with a third helix present during RNA binding in some cases. The C-terminal beta strand (4th strand) and final helix are hard to align and have been omitted in the SEED alignment. The LA proteins have a N terminus rrm which is included in the seed. There is a second region towards the C terminus that has some features of a rrm but does not appear to have the important structural core of a rrm. The LA proteins are one of the main autoantigens in Systemic lupus erythematosus (SLE), an autoimmune disease.

CD-Length = 71 residues, 95.8% aligned

Score = 47.8 bits (112), Expect = 1e-06

The cyclophilins are a conserved class of proteins that bind the immunosuppressive drug cyclosporin A (CsA) with high affinity. CsA blocks helper T-cell activation at a step between T-cell receptor stimulation and the transcriptional activation of cytokine genes.

Cyclophilins from many species possess peptidyl-prolyl cis-trans isomerase (PPIase) activity that is blocked by CsA and therefore may be relevant in CsA-mediated immunosuppression. Probing with the previously known cyclophilin cDNA under reduced stringencies, Price et al. (1991) identified a second cyclophilin gene, which encoded cyclophilin B (CYPB). The deduced protein was 64% identical to CYPA and was distinguished from it by a signal sequence that probably directs it to the endoplasmic reticulum (ER). CYPB showed even stronger similarity to yeast CYPB, which also has an ER-directed signal sequence. The signal sequence is removed from the protein upon expression in *E. coli*, and the processed protein possesses PPIase activity that is inhibited by CsA. Peddada et al. (1992) used the PCR technique to generate a unique probe complementary to the hydrophobic 5-prime end of the human cyclophilin B gene. Using this probe in an analysis of human/hamster hybrid somatic cell lines, they assigned the gene to chromosome 15. The human PIN1 gene encodes an essential nuclear peptidyl-prolyl cis/trans isomerase involved in the regulation of mitosis. PIN1 is a member of a new class of peptidyl-prolyl cis/trans isomerases that includes the *Escherichia coli* parvulin, yeast ESS1, and *Drosophila melanogaster* dodo gene products. Lu et al. (1996) described human PIN1 and showed that deletion of PIN1 from HeLa cells induces mitotic arrest, while HeLa cells overexpressing PIN1 arrest in the G2 phase. Campbell et al. (1997) identified a gene closely related to the gene encoding the essential nuclear peptidyl-prolyl cis/trans isomerase (PIN1) involved in the regulation of mitosis. The novel gene, called PIN1L by them, is 89% identical at the nucleotide level to the PIN1 transcript, but contains a shift in the reading frame. It encodes a 100-amino acid variant protein consisting of 63 amino acids homologous (90% identical) to PIN1 and contains the entire WW domain, fused to a 37-amino acid tail. By fluorescence in situ hybridization and somatic cell hybrid analysis, Campbell et al. (1997) mapped PIN1L to 1q31. They commented that the protein encoded by PIN1L may have some functional role or, alternatively, PIN1L may be a transcribed pseudogene. Campbell et al. (1997) found by analysis of human expressed sequence tags (ESTs) 2 different but closely related human transcripts, 1 of which corresponds to PIN1. Gene localization, using both fluorescence in situ hybridization and tritium-labeled probes, showed that each of the human transcripts hybridized to 1p31 and 19p13. Primers were designed to discriminate between the 2 transcripts, and PCR on DNA from hamster/human somatic cell hybrids retaining chromosomes 1 or 19 was used to map the human PIN1 gene to chromosome 19, and PIN1L, the closely related gene, to chromosome 1. Their results established that PIN1 is at 19p13 and PIN1L at 1p31. PCR was used to clone the coding region for PIN1L. The PIN1L cDNA is 89% identical at the nucleotide level to the PIN1

25

30

isomerase-like activities and physiological functions, or a fragment of such a nucleic acid.

The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

10 In the mutant or variant nucleic acids, and their complements, up to about 37 percent of the bases may be so changed.

The disclosed NOV34 protein of the invention includes the Cyclophilin-type peptidyl-prolyl cis-trans isomerase-like protein whose sequence is provided in Table 34B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 34B while still encoding a protein that maintains its Cyclophilin-type peptidyl-prolyl cis-trans isomerase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 24 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this Cyclophilin-type peptidyl-prolyl cis-trans isomerase-like protein (NOV34) may function as a member of a "Cyclophilin-type peptidyl-prolyl cis-trans isomerase family". Therefore, the NOV34 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV34 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the Cyclophilin-type peptidyl-prolyl cis-trans isomerase-like protein (NOV34) may be useful in gene therapy, and

the Cyclophilin-type peptidyl-prolyl cis-trans isomerase-like protein (NOV34) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from Cardiomyopathy, Atherosclerosis, Hypertension, Congenital heart defects, Aortic stenosis, Atrial septal defect (ASD), Atrioventricular (A-V) canal defect, Ductus arteriosus, Pulmonary stenosis, Subaortic stenosis, Ventricular septal defect (VSD), valve diseases, Tuberous sclerosis, Scleroderma, Obesity, Transplantation, Adrenoleukodystrophy, Congenital Adrenal Hyperplasia, Hemophilia, Hypercoagulation, Idiopathic thrombocytopenic purpura, Immunodeficiencies, Graft versus host, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, or other pathologies or conditions. The NOV34 nucleic acid encoding the Cyclophilin-type peptidyl-prolyl cis-trans isomerase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV34 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV34 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV34 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV35

A disclosed NOV35 nucleic acid of 1647 nucleotides (also referred to as CG57670-01) encoding a pyruvate kinase-like protein is shown in Table 35A. The start and stop codons are in bold letters.

Table 35A. NOV35 nucleotide sequence (SEQ ID NO:79).

GACCTCAGAAGCCATGTTGAAGCCCCATAGTGAAGCCAGGGCTGCCTTCATTACAGACCCAGCAGCTGCAC
GCAGCCATGGCTGACACATCTCTGGAGCACATGTGCTGCCTGGACACTGACTCGCCACCCATCACAGCCT
GGAGCACTGGCATCATCTGTACTATGGGCCAGCTTCTCCATTGCTAGAGATGCTGAAGAAAACGATTAA
GTCTGGAATTAATGTGGCTCATCTGAACTCTCATGGAGCCCATGAGTACCATACAGAGACCATCAAGAAC
GTCTGCACAGCCACGGAAGCTTTTGCTTCTGACTCCCTCTACCAGCCCATTTGCTGTGGCTCCAGACA
CTAAAGGACCTGAGATCCCAACTGGGCCCGTCAAGGGCAGCGGCACCTGCAGAGGTGGAGCTGAAGAAGGG
AGCCACTCTCAAGTTCACGCTGGATAATACCTACATGGAAGGGTAAAGAGAACATCCTGTGGCGGGAC
TACAAGAACATCTGCAAGGTGGTGGAGTGGGCTGCAAGATCTACGTGGATGATGGGCTAATTTCTCTCC
AAGTGAAGCAGAAGGATGCTCACTTTCTGGTGACAGAGGTGGAAAAATGGTGGCTCCTTGGGCAGCAAGAA
GAGTGTGAACCTTCTCTGGGGCTGCCGTGGACCTGTCTGCCATTGTGGAGAAGGACATCCAGGACCTGAAG
TTTGGGGGCGAGCAAGATGTCGATATGATGTTTTTCATCATTCTGCAAGACATCTGATGTCCATGAAG
TTAGGAAGGCTCTTGGGAGAGAAAGGAAAGAACAGCAAGATAACAGCAAAATTGAGAATCATGATGGGGG
TTGGAGGTTTGAATTAATCTGTGAGGCCAGCGATGGGATTATGGTAGCTCGTGGTGATCCACCACAAGCC
GTGAGATGGAGATTCTCTCAGGGAAGGCTTGCTGCTCAGAGGATGATGATTCGGTGGTGCAACCAAG
CTGGGAAGCCTGTCTATCTTTGCCACTCAGATGCTAGAGGATGTGATCAAGAAGCTCCGACCCCATTTGGGC
TGAGGGCAGTGGTGTGGCCAATGCAGTTCTTGGTGAAGCTGACTGCATCATGTCTGTCTGGAGAAACAGCC
AAAGGGAACATATCCTCTGGAGGCTGTGCACATGCAGCACCTGATTGCTGCTGAGGCAGAGGCCACCATCT
ACCACTTGCAATTAATTTGAGGAGTTCTGCCACTGGCACCCATTACAGTGACCCCGAGAAGCTACTGCA
CATGGGCACGTGGAGGCTCCTTCAAGTGTGTCAGTGGGGCCATATCGTCTCACAAGTCTGCCAGG
TGTGCCACCAGGTGGCCAGATACTGCCACCGTGCCCCCATGATTGTTGTGACATGGCATCCCCAGGCAG
CTCGCCAGGCCACCTGTACCGTGGTATCTTCCCTGTGCTGTGTAAGGACCCCATCCAGGAGCCCCAGGC
TGAGGATGTGGACCTCCGAGTGAACCTTGGCCATGAATGTTGGTAAGGCCCGAGGCTTCTTCAAGAAGGAT
GATGTGGTCATTGTGCTGACCTGGGGACACCTGGCCCTGGCTTCTCCACCACCTGTGTGTTATTCCTG
TGCTGTGATGAATCCAGAGCTCTTCTCCAGCCCT

5

10

Table 35B. Encoded NOV35 protein sequence (SEQ ID NO:78).

MLKPHSEARAAFIQTQQLHAAMADTFLEHMCCLDTSPPITAWSTGIICTMGPPASPLLEM LKKTIKSGINVAHLNLSHGAHEYHTTEITKNVRTATESFASDSILYQPIAVAPDTKGPEIPT GPVKGSGTAEVELKKKGATLKFTLDNTIMEKGKENILWRDYKINICKVVEVGSKIYVDDGLI SLQVKQKDAHFLVTEVGENSGSLGSKKSVNLPGAADVLSAMLEKDIQDLKFGGEQDVMFMF SSFICKTSDVHEVRKVLGEKGKNSKITSKIENHDGGWRPFDEILEASDGIMVARGDPPQAV EMEIPAGKVCLAQRMMIGWCNQAGKPVIFATQMLEDVIKKLHPTWAEGSGVANAVLVEAD CIMLSGETAKGNYPLEAVHMQHLIACEAEATIYHLQLFEEFCHLAPITSDPAEATAMGTV EASFKCCSGAIIIVLTKSARCAHQVARYCPRAPMIVVTHWPQAARQAHLYRGIFPVLCKDP IQEPQAEDVDLRVNLAMNVGKARGFFKKDDVVIVLTWGHPPGPGFSTTLCVIPVL
--

433/533 (81%) identity and 458/533 (85%) similarity with pir-id:S30038 pyruvate kinase (EC

Table 35D. Domain Analysis of NOV35

gnl|Pfam|pfam02887, PK_C, Pyruvate kinase, alpha/beta domain
 CD-Length = 116 residues, 99.1% aligned
 Score = 129 bits (324), Expect = 4e-31

Table 35E. Domain Analysis of NOV35

gnl|Pfam|pfam00224, PK, Pyruvate kinase, barrel domain. This domain of the is actually a small beta-barrel domain nested within a larger TIM barrel. The active site is found in a cleft between the two domains.
 CD-Length = 349 residues, 100.0% aligned
 Score = 422 bits (1084), Expect = 3e-119

Pyruvate kinase is also known as ATP:pyruvate phosphotransferase (EC 2.7.1.40). At least 3 molecular forms with pyruvate kinase activity are known (Bigley et al., 1968). The form that is deficient in a type of hemolytic anemia is the red cell variety, PK1. PK2 is found in kidney. PK3 is found in leukocytes, muscle, platelets, and brain but not in red cells or kidney. PK1 is found also in liver. A patient with red cell PK deficiency has been found to have abnormal liver enzyme also (Bunn, 1981); see Nakashima et al. (1977). During fetal development, PK3 changes to PK1 in the liver. PK1 is a tetramer composed of two dissimilar polypeptides of somewhat different molecular weight. It is an allosteric enzyme exhibiting cooperative binding for phosphoenolpyruvate and sensitivity to fructose-1,6-diphosphate. PK3 also is a tetrameric protein but, unlike PK1, all subunits are alike and, not unexpectedly, there is no cooperative behavior. The enzyme is insensitive to fructose-1,6-diphosphate. Patients with deficiency of red cell PK have normal PK2 and PK3. Tsutsumi et al. (1988) showed that pyruvate kinase occurs in 4 isozymic forms (L, R, M1, M2) and that these are encoded by 2 different genes, PKL and PKM. The L and R isozymes are generated from the PKL gene by differential splicing of RNA; the M1 and M2 forms are produced from the PKM gene by differential splicing. Studies of somatic cell hybrids showed that the PK3 and MPI loci are syntenic (Shows, 1972). By cell hybridization studies, Van Heyningen et al. (1975) found that the MPI and PK3 loci are on chromosome 15. Chern et al. (1977) narrowed the assignment to 15q22-qter. Tani et al. (1988) isolated and sequenced 2 overlapping clones covering the entire coding sequence of PKM2. By in situ hybridization they demonstrated that the gene is located at band 15q22. Northern blot analysis with RNA from a human hepatoma demonstrated that the M2-type PK was predominantly expressed in hepatoma cells, whereas L-type PK was preferentially expressed in the nontumor portion of the liver. Takenaka et al. (1991) reported that the gene that encodes both the M1 and the M2 isozymes is approximately 32 kb long and

comprises 12 exons and 11 introns. Exons 9 and 10 contain sequences specific for the M1 and M2 types, respectively, indicating that the human fetal and adult isozymes are produced from the same gene by alternative splicing.

5 The disclosed NOV35 nucleic acid of the invention encoding a pyruvate kinase-like protein includes the nucleic acid whose sequence is provided in Table 35A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 35A while still encoding a protein that maintains its pyruvate kinase-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are
10 complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or
15 derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 16 percent of the bases may be so changed.

The disclosed NOV35 protein of the invention includes the pyruvate kinase-like
20 protein whose sequence is provided in Table 35B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 35B while still encoding a protein that maintains its pyruvate kinase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 19 percent of the residues may be so changed.

25 The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this pyruvate kinase-like protein (NOV35) may function as a member of a "pyruvate kinase family". Therefore, the NOV35 nucleic acids and proteins identified here may be useful in potential therapeutic
30 applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene

delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV35 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the pyruvate kinase-like protein (NOV35) may be useful in gene therapy, and the pyruvate kinase-like protein (NOV35) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from Cardiomyopathy, Atherosclerosis, Hypertension, Congenital heart defects, Aortic stenosis, Atrial septal defect (ASD), Atrioventricular (A-V) canal defect, Ductus arteriosus, Pulmonary stenosis, Subaortic stenosis, Ventricular septal defect (VSD), valve diseases, Tuberous sclerosis, Scleroderma, Obesity, Transplantation, Adrenoleukodystrophy, Congenital Adrenal Hyperplasia, Hemophilia, Hypercoagulation, Idiopathic thrombocytopenic purpura, Immunodeficiencies, Graft versus host, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, or other pathologies or conditions. The NOV nucleic acid encoding the pyruvate kinase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV35 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV35 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV35 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV36

NOV36 includes two Cis/Trans Peptidyl Prolyl Isomerase-like proteins disclosed below. The disclosed sequences have been named NOV36a and NOV36b.

NOV36a

A disclosed NOV36a nucleic acid of 600 nucleotides (also referred to as CG57149-01) encoding a Cis/Trans Peptidyl Prolyl Isomerase-like protein is shown in Table 36A. The start and stop codons are in bold letters.

5

Table 36A. NOV36a nucleotide sequence (SEQ ID NO:81).

ACCAGGAGCCCTGTACTACCAGCCATGGTCAACCCACCATGTTCTTCAACATCGCCATCAACAGCGAGG CCTTGGGGCAGTCTCCTTCGAACGTGTTGCAGACAAGTTTCCAAAGACAGAAAACCTTCGTGCTCTGAG CACTGGAGAGAAAGGATTGGTTATAAGGGTTCCTGCTTTCACAGAATTATCTAGGGCTTTGTGTCAG GGTGGTGACTTTACATGCCATAATGGCACTGGTGGCAAGTCTGTCTACAGGGAGAAATTGATGATGAGA ACTTCATTCTGAAGCATAACAGGTCTGGCATCTTGTCATGAAGCATAACAGGTCTGGCATCTTGTCAT GGCAAATGCTGGACCCAAACAAACGATTCCAGATTTTCATCTGCACTGCCAAGACCGAGTGGTTGGAT GGCAAGCATGTGGTCTCTGGCAGGGTGAAAGAAGGCATCAAGATTGTGGAGGCCATGAAGCGCTATGGGT CCAAGAATGGCAAGAGCAGGAAGAAGATCACCCTGCTGACTGTGGACAACCTCTAATAAGTTTGACTTGT GTTTTATCTTAACCACCAGACCATTCTTTTGTAGCTCAG

In a search of public sequence databases, the NOV36a nucleic acid sequence, located on chromosome 10 has 288 of 327 bases (88%) identical to a gb:GENBANK-ID:HSCYCR|acc:Y00052.1 mRNA from Homo sapiens (Human mRNA for T-cell cyclophilin). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

10

The disclosed NOV36a polypeptide (SEQ ID NO:82) encoded by SEQ ID NO:81 has 173 amino acid residues and is presented in Table 36B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV36a has no signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.6000.

15

Table 36B. Encoded NOV36a protein sequence (SEQ ID NO:82).

MVNPTMFFNIAINSEALGHVSFELFADKFPKTENFRALSTGEKGFYKGSFHRILGLL CQGGDFTCHNGTGGKSVYREKFDDENFILKHTGPGILSMKHTGPGILSMANAGPNTNDSQ IFICTAKTEWLDGKHVVSGRVKEGIKIVEAMKRYGSKNGKSRKKITTADCGQL
--

A search of sequence databases reveals that the NOV36a amino acid sequence has 136 of 173 amino acid residues (78%) identical to, and 149 of 173 amino acid residues (86%) similar to, the 165 amino acid residue ptnr:pir-id:CSHUA protein from human (peptidylprolyl isomerase (EC 5.2.1.8) A). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

20

NOV36b

A disclosed NOV36b nucleic acid of 566 nucleotides (also referred to as CG57149-02) encoding a Cis/Trans Peptidyl Prolyl Isomerase-like protein is shown in Table 36A. The start and stop codons are in bold letters.

Table 36C. NOV36b nucleotide sequence (SEQ ID NO:83).

```
GTACTACCAGCCATGGTCAACCCACCATGTTCTTCAACATCGCCATCAACAGCGAGGCC
TTGGGGCACGTCCTCTCGAACTGTTTGAGACAAGTTTCCAAAGACAGAAAACTTTCGT
GCTCTGAGCACTGGAGAGAAAGGATTTGGTTATAAGGGTTCCTGCTTTCACAGAATTATT
CTAGGGGCTTTGTGTGTCAGGGTGGTGACTTACATGCCATAATGGCACTGGTGGCAAGTCT
GTCTACAGGGAGAAATTTGATGATGAGAACTTCATTCTGAAGCATACAGGTCCTGGCATC
TTGTCCATGAAGCATACAGGTCCTGGCATCTTGTCCATGGCAAATGCTGGACCCAACACA
AACGATTTCCAGATTTTCATCTGCACTGCCAAGACCGAGTGGTTGGATGGCAAGCATGTG
GTCTCTGGCAGGGTGAAAGAAGGCATCAAGATTGTGGAGGCCATGAAGCGCTATGGGTCC
AAGAATGGCAAGAGCAGGAAGAAGATCACCACCTGCTGACTGTGGACAACTCTAATAAGTT
TGACTTGTGTTTTATCTTAACCACCA
```

In a search of public sequence databases, the NOV36b nucleic acid sequence has 269 of 311 bases (86%) identical to a gb:GENBANK-ID:AF139893|acc:AF139893.1 mRNA from *Oryctolagus cuniculus* (*Oryctolagus cuniculus* cyclophilin 18 mRNA, complete cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV36b polypeptide (SEQ ID NO:84) encoded by SEQ ID NO:83 has 173 amino acid residues and is presented in Table 36B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV36b has no signal peptide and is likely to be localized in the cytoplasm with a certainty of 0.4500.

Table 36D. Encoded NOV36b protein sequence (SEQ ID NO:84).

```
MVNPTMFFNIAINSEALGHVS FELFADKFPKTENFRALSTGEKGFYKGS CFHRI ILGLL
CQGGDFTCHNGTGGKSVYREKFDDENFILKHTGPGILSMKHTGPGILSMANAGPNTNDSQ
IFICTAKTEWLDGKHVVS GRVKEGIKIVEAMKRYGSKNGKSRKKITTADCGQL
```

A search of sequence databases reveals that the NOV36b amino acid sequence has 136 of 173 amino acid residues (78%) identical to, and 149 of 173 amino acid residues (86%) similar to, the 165 amino acid residue ptnr:TREMBLNEW-ACC:AAH00689 protein from *Homo sapiens* (Human) (PEPTIDYLPROLYL ISOMERASE A (CYCLOPHILIN A)). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV36b is expressed in at least Epidermis, Lymphoid tissue. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

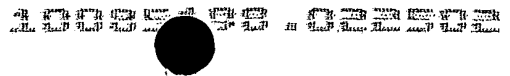
The disclosed NOV36a polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 36E.

Table 36E. BLAST results for NOV36a					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
<u>gi 12804335 gb AAH03026.1 AAH03026</u> (BC003026)	(protein for IMAGE:2823490) [Homo sapiens]	174	136/174 (78%)	149/174 (85%)	2e-70
<u>gi 4033689 sp P04374 CYPH BOVIN</u>	PEPTIDYL-PROLYL CIS-TRANS ISOMERASE A (PPIASE) (ROTAMASE) (CYCLOPHILIN A) (CYCLOSPORIN A- BINDING PROTEIN)	164	136/174 (78%)	149/174 (85%)	8e-70
<u>gi 10863927 ref NP_066953.1 </u> (NM_021130)	peptidylprolyl isomerase A (cyclophilin A) [Homo sapiens]	165	136/174 (78%)	149/174 (85%)	1e-69
<u>gi 68401 pir CSBOA B</u>	peptidylprolyl isomerase (EC 5.2.1.8) A - bovine	163	135/173 (78%)	148/173 (85%)	4e-69
<u>gi 13543666 gb AAH05982.1 AAH05982</u> (BC005982)	peptidylprolyl isomerase A (cyclophilin A) [Homo sapiens]	165	136/174 (78%)	149/174 (85%)	6e-69

5 Table 36F lists the domain descriptions from DOMAIN analysis results against NOV36a. This indicates that the NOV 36a sequence has properties similar to those of other proteins known to contain this domain.

Table 36F. Domain Analysis of NOV36	
<u>gnl Pfam pfam00160</u> , pro_isomerase, Cyclophilin type peptidyl-prolyl cis-trans isomerase	
CD-Length = 162 residues, 100.0% aligned	
Score = 196 bits (497), Expect = 1e-51	

10 The human parvulin Pin1 is a member of the peptidyl-prolyl cis-trans isomerase group of proteins, which modulate the assembly, folding, activity, and transport of essential cellular proteins. Pin1 is a mitotic regulator interacting with a range of proteins that are phosphorylated before cell division. In addition, an involvement of Pin1 in the tau-related neurodegenerative brain disorders has recently been shown. In this context, Pin1 becomes depleted from the
15 nucleus in Alzheimer's disease (AD) neurons when it is redirected to the large amounts of hyperphosphorylated tau associated with the neurofibrillary tangles. This depletion from the



nucleus may ultimately contribute to neuron cell death. The 131-amino acid residue parvulin-like human peptidyl-prolyl cis/trans isomerase (PPIase) hPar14 was shown to exhibit sequence similarity to the regulator enzyme for cell cycle transitions human hPin1, but specificity for catalyzing pSer(Thr)-Pro cis/trans isomerizations was lacking. That FK and CsA completely inhibit immune function without completely inhibiting CN suggests that the inhibition of immune function is not mediated by general CN inhibition but by inhibition of a subset of CN which is critical for lymphocyte activation.

The disclosed NOV36b nucleic acid of the invention encoding a Cis/Trans Peptidyl Prolyl Isomerase-like protein includes the nucleic acid whose sequence is provided in Table 36A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 36A while still encoding a protein that maintains its Cis/Trans Peptidyl Prolyl Isomerase-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 14 percent of the bases may be so changed.

The disclosed NOV36b protein of the invention includes the Cis/Trans Peptidyl Prolyl Isomerase-like protein whose sequence is provided in Table 36B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 36B while still encoding a protein that maintains its Cis/Trans Peptidyl Prolyl Isomerase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 22 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this Cis/Trans Peptidyl Prolyl Isomerase-like protein (NOV36) may function as a member of a "Cis/Trans Peptidyl Prolyl Isomerase family". Therefore, the NOV36 nucleic acids and proteins identified here

5

20

30

NOV37

A disclosed NOV37 nucleic acid of 660 nucleotides (also referred to as CG57151-01) encoding a Cis/Trans Peptidyl Prolyl Isomerase-like protein is shown in Table 37A. The start and stop codons are in bold letters.

5

Table 37A. NOV37 nucleotide sequence (SEQ ID NO:85).
ACTAGTCATTCTTCCAGTAGCTA ATGA AGCTGACTTTTAAAAAGAAGGCTGTGAGCTTTGCAGATGCTG CTGCCGCCAGGGCCCCCTGCTTCCAGCCATGGTCAACCCACCATGTTTTCACATTGCTGTCGATGG CGAGCCCTTGGGCTGTGTCTCCTTCGAGGTAGAGCTGTTTGCAGACAAGGTTCCAAAGACAGCAGAAAAT TTCCATGCTCTGAGCACTGGAGAAAAAGGATTGGTTATAAGGGTTCCTGCTTTCACAGAATTATTCAG GGTTACGTGTCAGAGTGGTGACTTCACACGCCATGGTGGCAAGTCCATCTGCAGGGAGAAATTTGATGA CAAGAACTTCATCCTGAAGCATACGGGTCCTGGCATCTTGTCCATGGCAAATGCTGGACCCAGCGTGAAC GTTCCCGAGTTTATCTGCCCTGCCAAGACAGAGTGGTTGGATTGCAAGCATGTGGTCTTTGGCAAGG TGAAGATGGCATGAATATTGTGGAGGTCATGGAGCACTTGGGGTCCAAGAATGGCAAGATCAGCAAGAA GATCACCATTGCTGACTGGACAACGCAATAAATTGACGGGTGTTCTCTTAAAAAAAAAAAAAATA CTGT GAC AGACCAAGGTAAATTGTTTGA

In a search of public sequence databases, the NOV37 nucleic acid sequence, located on chromosome 10 has 492 of 581 bases (85%) identical to a (HSCPH192|acc: X52857.1) cyclophilin-related processed pseudogene mRNA from human. Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

10

The disclosed NOV37 polypeptide (SEQ ID NO:86) encoded by SEQ ID NO:85 has 203 amino acid residues and is presented in Table 37B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV37 has no signal peptide and is likely to be localized in the cytoplasm with a certainty of 0.6400.

Table 37B. Encoded NOV37 protein sequence (SEQ ID NO:86).
MKLTFKKKAVSFADAAAAQGPLLPAAMVNPTMFFHIAVDGEPLGCVSFEVELFADKVPKTA ENFHALSTGEKGFGYKGSFHRHII PGFTCSGDFTRHGGKSI CREKFDDKNFILKHTGPG ILSMANAGPSVNVSQFFICPAKTEWLDCKHVVFGKVKDGMNIVEVMEHLGSKNGKISKKI TIADWTTAINLTGVSLKKKKKIL

15

A search of sequence databases reveals that the NOV37 amino acid sequence has 136 of 160 amino acid residues (85%) identical to, and 142 of 160 amino acid residues (88%) similar to, the 165 amino acid residue ptnr:pir-id:CSHUA protein from human (peptidylprolyl isomerase (EC 5.2.1.8) A). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

20

NOV37 is expressed in at least Bone Marrow, Brain, Cartilage, Cochlea, Colon, Epidermis, Kidney, Lung, Mammary gland/Breast, Ovary, Pancreas, Prostate, Stomach, Testis, Thymus, Umbilical Vein, Uterus, Vulva, Whole Organism. Expression information

was derived from the tissue sources of the sequences that were included in the derivation of the sequence of CG57151_01. The sequence is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: pir-id:CSHUA peptidylprolyl isomerase) a closely related peptidylprolyl isomerase homolog in species human: Kidney, Lung. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV37 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 37C.

10

Table 37C. BLAST results for NOV37					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 12804335 gb AAH03026.1 AAH03026 (BC003026)	protein for IMAGE:2823490) [Homo sapiens]	174	139/171 (81%)	147/171 (85%)	E3-71
gi 4033689 sp P04374 CYPH BOVIN	PEPTIDYL-PROLYL CIS-TRANS ISOMERASE A (PPIASE) (ROTAMASE) (CYCLOPHILIN A) (CYCLOSPORIN A-BINDING PROTEIN)	164	137/162 (84%)	142/162 (87%)	4e-70
gi 10863927 ref NP_066953.1 (NM_021130)	peptidylprolyl isomerase A (cyclophilin A) [Homo sapiens]	165	136/162 (83%)	142/162 (86%)	1e-69
gi 68401 pir CSBOA B	peptidylprolyl isomerase (EC 5.2.1.8) A - bovine	163	136/161 (84%)	141/161 (87%)	2e-69
gi 13937981 gb AAH07104.1 AAH07104 (BC007104	peptidylprolyl isomerase A (cyclophilin A) [Homo sapiens]	165	136/162 (83%)	141/162 (86%)	2e-69

Table 37D lists the domain descriptions from DOMAIN analysis results against NOV37. This indicates that the NOV37 sequence has properties similar to those of other proteins known to contain this domain.

15

Table 37D. Domain Analysis of NOV37

gnl|Pfam|pfam00160, pro_isomerase, Cyclophilin type peptidyl-prolyl
cis-trans isomerase
CD-Length = 162 residues, 97.5% aligned
Score = 191 bits (484), Expect = 5e-50

The disclosed NOV37 nucleic acid of the invention encoding a Cis/Trans Peptidyl Prolyl Isomerase-like protein includes the nucleic acid whose sequence is provided in Table 37A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 37A while still encoding a protein that maintains its Cis/Trans Peptidyl Prolyl Isomerase-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 15 percent of the bases may be so changed.

The disclosed NOV37 protein of the invention includes the Cis/Trans Peptidyl Prolyl Isomerase-like protein whose sequence is provided in Table 37B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 37B while still encoding a protein that maintains its Cis/Trans Peptidyl Prolyl Isomerase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 15 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this Cis/Trans Peptidyl Prolyl Isomerase-like protein (NOV37) may function as a member of a "Cis/Trans Peptidyl Prolyl Isomerase family". Therefore, the NOV37 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various

pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue
5 regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV37 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the Cis/Trans Peptidyl
10 Prolyl Isomerase-like protein (NOV37) may be useful in gene therapy, and the Cis/Trans Peptidyl Prolyl Isomerase-like protein (NOV37) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from CNS disorders, brain disorders including epilepsy, eating disorders, schizophrenia, ADD; cancer; heart disease; inflammation
15 and autoimmune disorders including Crohn's disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, blood disorders; psoriasis colon cancer, leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, pancreatic disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer, or other pathologies
20 or conditions. The NOV37 nucleic acid encoding the Cis/Trans Peptidyl Prolyl Isomerase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV37 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV37 substances for use in
25 therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV37 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in
30 understanding of pathology of the disease and development of new drug targets for various disorders.

NOV38

A disclosed NOV38 nucleic acid of 600 nucleotides (also referred to as CG57153-01) encoding a Cis/Trans Peptidyl Prolyl Isomerase-like protein is shown in Table 38A. The start and stop codons are in bold letters.

5

Table 38A. NOV38 nucleotide sequence (SEQ ID NO:87).

AATTTATTGGTTTGTGTTTAAAAATTTGTTGGCACTTTGCAGATGCCACTTCCACTGATGTCACCAC TGCCAGTGATGGTCACCTCCACCTTGTTCTTTAACTTTGTAGTCAACGGTGAGCACTTGGGCCATGTCTC CTTCCAGCTGTTTGCAAAGAAAGTTCCAAAGACAGCAGAAAATGTTCAATTTGTGAGCACTGGAGAGAAA GGATTTGGCTATAAGTGTTCCCTGTTTTCACAGAATTATTCAGGGTTTATATGCCAGAGTGGTGACTTCA CATGTCATGATGACACTGGCACAAAGTCCAACACTGGGAGAAGTCTGATGATGATAACTCCATCCTGAA GCATACAAGACCTGGCACCTTGTCATGGCAAATACTGGACGCTACACAAATGGTTTCCAGTTTTTCATC TGCACTGCCAAACTGTGTGGTTGGGTGGCAAGAGTGCACTCTTGGCAAGACAAAAGAGGGCTTGAATA TCTTGAAGCCATGGCGCACTTTGCTTCTGGAATGGCAAACCAGAAAGAAGACCACGATTGACAACCTG TGGACAACCTCCAATAAAATTTAACTTATGTTTTGTTTAAAC

In a search of public sequence databases, the NOV38 nucleic acid sequence, located on chromosome 16 has 456 of 564 bases (80%) identical to a gb:GENBANK-

ID:AK026569|acc:AK026569.1 mRNA from Homo sapiens (Homo sapiens cDNA: FLJ22916
10 fis, clone KAT06406, highly similar to HSCYCR Human mRNA for T-cell cyclophilin.

Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV38 polypeptide (SEQ ID NO:88) encoded by SEQ ID NO:87 has 176 amino acid residues and is presented in Table 38B using the one-letter amino acid code.

Signal P, Psort and/or Hydropathy results predict that NOV38 has no signal peptide and is

15 likely to be localized in the cytoplasm with a certainty of 0.5500.

Table 38B. Encoded NOV38 protein sequence (SEQ ID NO:88).

MPLPLMSPLPVMVTSTLFFNFVNGEHLGHVSFQLFAKKVPKTAENVHVFVSTGEKGFYK CSCFHRIIPGFICQSGDFTCHDDTGTKSNYWEKSDDDNSILKHTRPGTLSMANTGRYTNG FQFFICTAKTVWLGGKSAVFGKTK EGLNILEAMAHFAFWNGKTRKKTIDNCGQLQ

A search of sequence databases reveals that the NOV38 amino acid sequence has 113
of 164 amino acid residues (68%) identical to, and 127 of 164 amino acid residues (77%)
20 similar to, the 164 amino acid residue ptnr:SPTREMBL-ACC:Q9TTC6 protein from
Oryctolagus cuniculus (Rabbit) (CYCLOPHILIN 18). Public amino acid databases include the
GenBank databases, SwissProt, PDB and PIR.

NOV38 is expressed in at least Bone Marrow, Brain, Cartilage, Cochlea, Colon,
Epidermis, Kidney, Lung, Mammary gland/Breast, Ovary, Pancreas, Prostate, Stomach,
25 Testis, Thymus, Umbilical Vein, Uterus, Vulva, and Whole Organism. Expression information

was derived from the tissue sources of the sequences that were included in the derivation of the sequence of CG57153_01. The sequence is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:AK026569|acc:AK026569.1) a closely related Homo sapiens cDNA: FLJ22916 fis, clone KAT06406, highly similar to HSCYCR Human mRNA for T-cell cyclophilin homolog in species Homo sapiens. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV38 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 38C.

Table 38C. BLAST results for NOV38					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 12804335 gb AAH03026.1 AAH03026 (BC003026)	(protein for IMAGE:2823490) [Homo sapiens]	174	115/168 (68%)	128/168 (75%)	6e-58
gi 13937981 gb AAH07104.1 AAH07104 (BC007104)	peptidylprolyl isomerase A (cyclophilin A) [Homo sapiens]	165	114/165 (69%)	127/165 (76%)	2e-57
gi 10863927 ref NP_066953.1 (NM_021130)	peptidylprolyl isomerase A (cyclophilin A) [Homo sapiens]	165	114/165 (69%)	127/165 (76%)	4e-57
gi 4033689 sp P04374 CYPH BOVIN	PEPTIDYL-PROLYL CIS-TRANS ISOMERASE A (PPIASE) (ROTAMASE) (CYCLOPHILIN A) (CYCLOSPORIN A-BINDING PROTEIN)	154	114/164 (69%)	126/164 (76%)	4e-57
gi 6651171 gb AAF22215.1 AF139893.1 (AF139893)	cyclophilin 18 [Oryctolagus cuniculus]	154	113/164 (68%)	127/164 (76%)	7e-57

Table 38D lists the domain descriptions from DOMAIN analysis results against NOV38. This indicates that the NOV38 sequence has properties similar to those of other proteins known to contain this domain.

Table 38D. Domain Analysis of NOV 38

gnl|Pfam|pfam00160, pro_isomerase, Cyclophilin type peptidyl-prolyl
cis-trans isomerase.

CD-Length = 162 residues, 100.0% aligned

Score = 171 bits (433), Expect = 3e-44

The disclosed NOV38 nucleic acid of the invention encoding a Cis/Trans Peptidyl Prolyl Isomerase-like protein includes the nucleic acid whose sequence is provided in Table 38A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 38A while still encoding a protein that maintains its Cis/Trans Peptidyl Prolyl Isomerase-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 20 percent of the bases may be so changed.

The disclosed NOV38 protein of the invention includes the Cis/Trans Peptidyl Prolyl Isomerase-like protein whose sequence is provided in Table 38B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 38B while still encoding a protein that maintains its Cis/Trans Peptidyl Prolyl Isomerase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 32 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this Cis/Trans Peptidyl Prolyl Isomerase-like protein (NOV38) may function as a member of a "Cis/Trans Peptidyl Prolyl Isomerase family". Therefore, the NOV38 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this

invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV38 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the Cis/Trans Peptidyl Prolyl Isomerase-like protein (NOV38) may be useful in gene therapy, and the Cis/Trans Peptidyl Prolyl Isomerase-like protein (NOV38) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from CNS disorders, brain disorders including epilepsy, eating disorders, schizophrenia, ADD; cancer; heart disease; inflammation and autoimmune disorders including Crohn's disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, blood disorders; psoriasis colon cancer, leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, pancreatic disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer, or other pathologies or conditions. The NOV38 nucleic acid encoding the Cis/Trans Peptidyl Prolyl Isomerase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV38 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV38 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV38 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV39

A disclosed NOV39 nucleic acid of 600 nucleotides (also referred to as CG57155-01) encoding a Cis/Trans Peptidyl Prolyl Isomerase-like protein is shown in Table 39A.

Table 39A. NOV39 nucleotide sequence (SEQ ID NO:89).

CTGAAAACTTTCTTGTAAGACTTTGACCTGCATTATATGATTCTCCTTAATCTTCACAGCATGATTTTCGT
GTTTTTGGACATTTCTATACGTGGATGTGTGTCCCAAACATGTAAAAATTTTCAGGTCTTGTGCACAGG
AAAAGACGGGTTTTCTCAACGTGGCATAAGACTACATTACAAAATTTCCATTTTTCATCGAATAGTACAG
AATGGCTGAGGATAAAGGGGATATAGTCTATGAAAAGGAGATATCCGAGAGTCGATTTATGGTCCAA
CATTTGAAGATGAAAACTTTTCAGTTTCTCTATAATAAAGAGGAGTACTTGGAAATGGCCAACAAGGCCG
TCACAGCAACGGGTCACAATTCTATATCACACTGCAAGCAACTCCTTATCTAGATAGAAAATTTGTGGCT
TTTGGGTATGTATATTGTAGATCTATTTATATAATATTCACACTGGTAGTAAAAAGCCCAGAGAAGTA
TGTGCAAGAACTAACAGTATGTGGTTGTGGCGCTAGTTTTTCAAAGGAAGAAGTAGTCAAATGCTGTAA
CAAGGACAACCTCATCTTGAACACTTACGCGATGGTGTGT

In a search of public sequence databases, the NOV39 nucleic acid sequence, located on chromosome 6 has 257 of 397 bases (64%) identical to a gb:GENBANK-

5 ID:AF043642|acc:AF043642.1 mRNA from *Rattus norvegicus* (*Rattus norvegicus* matrix
cyclophilin (matrin-cyp) mRNA, complete cds). Public nucleotide databases include all
GenBank databases and the GeneSeq patent database.

The disclosed NOV39 polypeptide (SEQ ID NO:90) encoded by SEQ ID NO:89 has 180 amino acid residues and is presented in Table 39B using the one-letter amino acid code.

10 Signal P, Psort and/or Hydropathy results predict that NOV39 has a signal peptide and is likely to be localized extracellularly with a certainty of 0.5128. The most likely cleavage site for a NOV39 peptide is between amino acids 19 and 20.

Table 39B. Encoded NOV39 protein sequence (SEQ ID NO:90).

MILLNLHSMISCFWTFLYCDVCPKTKNFQVLCTGKAGFSQRGIRLHYKNSIFHRIVQNGWIQGGDIVYG
KGDNGESIYGPTFEDENFSVPHNKRGLGMANKGRHSNGSQFYITLQATPYLDRKFVAFGYVYCRSIYII
FTPGSKKAQRSCKKLTVCGGRSFSKEEVVKCCNKDNSS

A search of sequence databases reveals that the NOV39 amino acid sequence has 70 of 87 amino acid residues (80%) identical to, and 80 of 87 amino acid residues (91%) similar to, the 278 amino acid residue ptnr:TREMBLNEW-ACC:BAB29003 protein from *Mus musculus* (Mouse) (ADULT MALE HIPPOCAMPUS CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY, CLONE:2900084F20, FULL INSERT SEQUENCE). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV39 is expressed in at least : Kidney, Lung, Lymphoid tissue, Mammary gland/Breast, Oviduct/Uterine Tube/Fallopian tube, Testis, Whole Organism. Expression information was derived from the tissue sources of the sequences that were included in the derivation of the sequence of CG57155_01. The sequence is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:AF043642|acc:AF043642.1) a closely related *Rattus norvegicus* matrix cyclophilin

(matrin-cyp) mRNA, complete cds homolog in species Rattus norvegicus: Kidney, Lung, Lymphoid tissue. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

5 The disclosed NOV39 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table39C.

Table 39C. BLAST results for NOV39					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 12851324 dbj BAB29003.1 (AK013818)	similar to NK-TUMOR RECOGNITION PROTEIN (NATURAL-KILLER CELLS CYCLOPHILIN-RELATED PROTEIN) (NK-TR PROTEIN) [Mus musculus]	278	70/87 (80%)	80/87 (91%)	2e-41
gi 6841028 gb AAF28867.1 (AF121134)	cyclophilin [Schistosoma mansoni]	181	65/120 (54%)	89/120 (74%)	3e-35
gi 13929124 ref NP113981.1 (NM_031793)	matrin cyclophilin (matrin-cyp) [Rattus norvegicus]	752	64/117 (54%)	85/117 (71%)	1e-31
gi 6754858 ref NP_035048.1 (NM_010918)	natural killer tumor recognition [Mus musculus]	1482	63/117 (53%)	82/117 (69%)	2e-31
gi 8039799 sp P30415 [NKR MOUSE]	NK-tumor recognition protein (Natural-killer cells; cyclophilin-related protein) (NK-TR protein)	1453	63/117 (53%)	82/117 (69%)	2e-31

10 Table 39D lists the domain descriptions from DOMAIN analysis results against NOV39. This indicates that the NOV39 sequence has properties similar to those of other proteins known to contain this domain.

Table 39E. Domain Analysis of NOV39
gnl Pfam pfam00160 , pro_isomerase, Cyclophilin type peptidyl-prolyl cis-trans isomerase
CD-Length = 162 residues, 67.9% aligned
Score = 141 bits (356), Expect = 3e-35

15 The disclosed NOV39 nucleic acid of the invention encoding a Cis/Trans Peptidyl Prolyl Isomerase-like protein includes the nucleic acid whose sequence is provided in Table

10

20

 $(F_{ab})_2$

30

therap

and disorders as indicated below. For example, a cDNA encoding the Cis/Trans Peptidyl Prolyl Isomerase-like protein (NOV39) may be useful in gene therapy, and the Cis/Trans Peptidyl Prolyl Isomerase-like protein (NOV39) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from CNS disorders, brain disorders including epilepsy, eating disorders, schizophrenia, ADD; cancer; heart disease; inflammation and autoimmune disorders including Crohn's disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, blood disorders; psoriasis colon cancer, leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, pancreatic disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer, or other pathologies or conditions. The NOV39 nucleic acid encoding the Cis/Trans Peptidyl Prolyl Isomerase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV39 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV39 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV39 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV40

A disclosed NOV40 nucleic acid of 572 nucleotides (also referred to as CG57157-01) encoding a Cis/Trans Peptidyl Prolyl Isomerase-like protein is shown in Table 40A. The start and stop codons are in bold letters.

Table 40A. NOV40 nucleotide sequence (SEQ ID NO:91).

TTTGTAGTCATCGCTGCCACCTGAAGCCACCTGCCTCTAGCCATGGTCAACGCCCCACTGTGTTCTTTTG
ACATCATTGTTGATGGTAACCTCTTGGCCCATGCAGCTCCTTCGAGCTGTTGCGGACAAAGTTCCAAA
AACAGTGGAAGAACTTTCGTGCACTGAGCACTGGAGGAAAAGGATTTGGTTATAAGGGTTCTCTGCTTTCAC
AGAATTATTCAGGGTTTATTTTATCTGCCAGAGTGCTGACTTCACACACCATAATAATGCCCCAGTCCA
TCTACCAGGAGAAATTTGATGATGAGAACTTCATCTTGAAGCACACAGGTCTGGCATCTTGTCATGGC
AAATGCTGGCCCGGACACAAATGGTTCCAGTTTTCACCTGTGTGGCCAAGACTGAGTGGCTGGATGGC
AAGCACAAAGGTCTTTGGCAAAGTGAGAAGAGGGGTGAATATCATGGAAGCCATGGAGTGTCTGCGGTCCG
GGAATGGTGAGACTGGCAAGAAGATCACCCTGCCAAGTGCAGCAACTCTAATCAATCTGCTTGTGTTT

GATCTTAACCAC

In a search of public sequence databases, the NOV40 nucleic acid sequence, located on chromosome 17 has 446 of 535 bases (83%) identical to a gb:GENBANK-

ID:HSCYCR|acc:Y00052.1 mRNA from Homo sapiens (Human mRNA for T-cell

cyclophilin). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV40 polypeptide (SEQ ID NO:92) encoded by SEQ ID NO:91 has 166 amino acid residues and is presented in Table 40B using the one-letter amino acid code.

Signal P, Psort and/or Hydropathy results predict that NOV40 has no signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.6000.

Table 40B. Encoded NOV40 protein sequence (SEQ ID NO:92).

MVNAPLCSFDIIVDGNISFGPCSSFELFADKVPKTVENFRALSTGGKGFYKGSCHFRIIP
GFILSARVLTSHITIMPQSIYQEKFDENFILKHTGPGILSMANAGPDNNGSQFFTCVAK
TEWLDGKHKVFVGKVRGVNIMEAMECSGSGNGETGKKITTANCGQL

A search of sequence databases reveals that the NOV40 amino acid sequence has 122 of 166 amino acid residues (73%) identical to, and 131 of 166 amino acid residues (78%)

similar to, the 165 amino acid residue ptnr:pir-id:CSHUA protein from human (peptidylprolyl isomerase (EC 5.2.1.8) A). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV40 is expressed in at least Brain. Expression information was derived from the tissue sources of the sequences that were included in the derivation of the sequence of CG57157_01. The sequence is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:HSCYCR|acc:Y00052.1) a closely related Human mRNA for T-cell cyclophilin homolog in species Homo sapiens: Brain. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources,

Literature sources, and/or RACE sources.

The disclosed NOV40 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 40C .

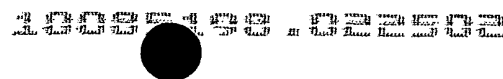
Table 40C. BLAST results for NOV40					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 14743515 ref XP_017224.2 (XM_017224)	hypothetical protein XP_017224 [Homo sapiens]	165	121/166 (72%)	130/166 (77%)	e-59
gi 12804335 gb AAH03026.1 AAH03026 (BC003026)	(protein for IMAGE:2823490) [Homo sapiens]	174	122/166 (73%)	131/166 (78%)	e-59
gi 4033689 sp P04374 CYPH_BOVIN	PEPTIDYL-PROLYL CIS-TRANS ISOMERASE A (PPIASE) (ROTAMASE) (CYCLOPHILIN A) (CYCLOSPORIN A- BINDING PROTEIN)	164	122/166 (73%)	131/166 (78%)	4e-59
gi 10863927 ref NP_066953.1 (NM_021130)	peptidylprolyl isomerase A (cyclophilin A) [Homo sapiens]	165	122/166 (73%)	131/166 (78%)	5e-59
gi 68401 pir CSBOA B	peptidylprolyl isomerase (EC 5.2.1.8) A - bovine	163	119/162 (73%)	128/162 (78%)	9e-59

Table 40D lists the domain descriptions from DOMAIN analysis results against NOV40. This indicates that the NOV40 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 40D. Domain Analysis of NOV40
gnl Pfam pfam00160, pro_isomerase, Cyclophilin type peptidyl-prolyl cis-trans isomerase
CD-Length = 162 residues, 99.4% aligned
Score = 185 bits (469), Expect = 2e-48

The disclosed NOV40 nucleic acid of the invention encoding a Cis/Trans Peptidyl Prolyl Isomerase-like protein includes the nucleic acid whose sequence is provided in Table 40A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 40A while still encoding a protein that maintains its Cis/Trans Peptidyl Prolyl Isomerase-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way



of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 17 percent of the bases may be so changed.

The disclosed NOV40 protein of the invention includes the Cis/Trans Peptidyl Prolyl Isomerase-like protein whose sequence is provided in Table 40B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 40B while still encoding a protein that maintains its Cis/Trans Peptidyl Prolyl Isomerase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 27 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this Cis/Trans Peptidyl Prolyl Isomerase-like protein (NOV40) may function as a member of a "Cis/Trans Peptidyl Prolyl Isomerase family". Therefore, the NOV40 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV40 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the Cis/Trans Peptidyl Prolyl Isomerase-like protein (NOV40) may be useful in gene therapy, and the Cis/Trans Peptidyl Prolyl Isomerase-like protein (NOV40) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from CNS disorders, brain disorders including epilepsy, eating disorders, schizophrenia, ADD; cancer; heart disease; inflammation and autoimmune disorders including Crohn's disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, blood disorders; psoriasis colon cancer, leukemia

AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, pancreatic disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer, or other pathologies or conditions. The NOV40 nucleic acid encoding the Cis/Trans Peptidyl Prolyl Isomerase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV40 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV40 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV40 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV41

A disclosed NOV41 nucleic acid of 525 nucleotides (also referred to as CG57159-01) encoding a Cis/Trans Peptidyl Prolyl Isomerase-like protein is shown in Table 41A. The start and stop codons are in bold letters.

Table 41A. NOV41 nucleotide sequence (SEQ ID NO:93).

GCCCAGAACTCCCTGCCACCAGCCATGGCCAACCCCACTGTGTTCTTCAACATTGCAATTGATAGTGAGT
CCTTGGGCTGCATCTCCTTCAAGCTATTTGCAGACAAAGTTCTAAAGATGGAAGAAAATTTTGTGCTCT
GAACACTGGAGAGAAAGTATTTGGTGATAAATGTCCCTGCTTTACAGAATTATTCGGGGGTGTGTCAG
GGTGGTGACTTCACACACCATAATGGCACTGGTGGCAAGTCCCTCTACAGCAAGGAATTTGATGATGAGA
ACTTCATCCTAAAGCATAACAGCTCCTGGCGTCTGTCCACGGCAAATGCTGGACCCACCACAAATGGTTC
CCAGTTTTTCTTCTGTACTGCCAAGACAGAGGATGGACAGCATGTGGTCTTTGGCAAGGTGAAAGATGGC
ATGAGTATTGTGGAAGCCCTGGAACGCTCTGGGTCCAGGAATGGTAAGACCAGCAAGAAGATCACAGCTG
CTGACTGTGGACAACCTCTAATAAATTTGATTGTTT

In a search of public sequence databases, the NOV41 nucleic acid sequence, located on chromosome 11 has 442 of 515 bases (85%) identical to a gb:GENBANK-ID:HSCYCR|acc:Y00052.1 mRNA from Homo sapiens (Human mRNA for T-cell cyclophilin). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV41 polypeptide (SEQ ID NO:94) encoded by SEQ ID NO:93 has 161 amino acid residues and is presented in Table 41B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV41 has no signal peptide and is likely to be localized in the plasma membrane with a certainty of 0.600.

5

Table 41B. Encoded NOV41 protein sequence (SEQ ID NO:94).
MANPTVFFNIAIDSESLGCISFKLFADKVLKMEENFCALNTGEKVFGDKCPCFYRIIPGV CQGGDFTHHNGTGGKSLYSKEFDDENFILKHTAPGVLSTANAGPTTNGSQFFCTAKTED GQHVVFQKVKDGMSIVEALERSGSRNGKTSKKITAADCGQL

A search of sequence databases reveals that the NOV41 amino acid sequence has 125 of 164 amino acid residues (76%) identical to, and 141 of 164 amino acid residues (85%) similar to, the 165 amino acid residue ptnr:pir-id:CSHUA protein from human (peptidylprolyl isomerase (EC 5.2.1.8) A). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV41 is expressed in at least Heart, Placenta, Stomach, Whole Organism. Expression information was derived from the tissue sources of the sequences that were included in the derivation of the sequence of CG57159_01. The sequence is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:HSCYCR|acc:Y00052.1) a closely related Human mRNA for T-cell cyclophilin homolog in species Homo sapiens : signet-ring cell carcinoma cell_line. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV41 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 41C.

Table 41C. BLAST results for NOV41					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
<u>gi 12804335 gb AAH03026.1 AAH03026</u> (BC003026)	(protein for IMAGE:2823490) [Homo sapiens]	174	125/164 (76%)	141/164 (85%)	3e-53

<u>gi 4033689 sp P04374 CYPH BOVIN</u>	PEPTIDYL-PROLYL CIS-TRANS ISOMERASE A (PPIASE) (ROTAMASE) (CYCLOPHILIN A) (CYCLOSPORIN A- BINDING PROTEIN)	164	125/164 (76%)	141/164 (85%)	9e-53
<u>gi 13937981 gb AAH07104.1 AAH07104 (BC007104)</u>	peptidylprolyl isomerase A (cyclophilin A) [Homo sapiens]	165	125/164 (76%)	141/164 (85%)	9e-53
<u>gi 10863927 ref NP_066953.1 (NM_021130)</u>	peptidylprolyl isomerase A (cyclophilin A) [Homo sapiens]	165	125/164 (76%)	141/164 (85%)	1e-52
<u>gi 68401 pir CSBOA B</u>	peptidylprolyl isomerase (EC 5.2.1.8) A - bovine	163	124/162 (76%)	140/162 (85%)	4e-52

Table 41D lists the domain descriptions from DOMAIN analysis results against NOV41. This indicates that the NOV41 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 41D. Domain Analysis of NOV41	
<u>gnl Pfam pfam00160, pro_isomerase, Cyclophilin type peptidyl-prolyl cis-trans isomerase</u>	
CD-Length = 162 residues, 100.0% aligned	
Score = 177 bits (448), Expect = 5e-46	

The disclosed NOV41 nucleic acid of the invention encoding a Cis/Trans Peptidyl Prolyl Isomerase-like protein includes the nucleic acid whose sequence is provided in Table 41A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of
10 whose bases may be changed from the corresponding base shown in Table 41A while still encoding a protein that maintains its Cis/Trans Peptidyl Prolyl Isomerase-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The
15 invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as
20 antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or

variant nucleic acids, and their complements, up to about 15 percent of the bases may be so changed.

The disclosed NOV41 protein of the invention includes the Cis/Trans Peptidyl Prolyl Isomerase-like protein whose sequence is provided in Table 41B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 41B while still encoding a protein that maintains its Cis/Trans Peptidyl Prolyl Isomerase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 24 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this Cis/Trans Peptidyl Prolyl Isomerase-like protein (NOV41) may function as a member of a “Cis/Trans Peptidyl Prolyl Isomerase family”. Therefore, the NOV41 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV41 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the Cis/Trans Peptidyl Prolyl Isomerase-like protein (NOV41) may be useful in gene therapy, and the Cis/Trans Peptidyl Prolyl Isomerase-like protein (NOV41) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from CNS disorders, brain disorders including epilepsy, eating disorders, schizophrenia, ADD; cancer; heart disease; inflammation and autoimmune disorders including Crohn's disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, blood disorders; psoriasis colon cancer, leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, pancreatic disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer, or other pathologies or conditions. The NOV41 nucleic acid encoding the Cis/Trans Peptidyl Prolyl Isomerase-like

protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV41 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV41 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV41 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV42

NOV42 includes two Cis/Trans Peptidyl Prolyl Isomerase-like proteins disclosed below. The disclosed sequences have been named NOV42a and NOV42b.

NOV42a

A disclosed NOV42a nucleic acid of 720 nucleotides (also referred to as CG57226-01) encoding a Cis/Trans Peptidyl Prolyl Isomerase -like protein is shown in Table 42A. The start and stop codons are in bold letters.

Table 42A. NOV42a nucleotide sequence (SEQ ID NO:95).

CATCAGGAAAATGCAAATCAAACCACAACGAGATATCATGTCACACCAATTAGGATGGCCACTATTAAAA
ACATAAAATTAATAAGCATTGGCAAGGATGTAGAAATTAGAACACCTGTGCACTGTTGGTGGGAATATAA
AATGATGCAGCTGGCTTTGCAGACACTGCTGTCCCCAACACCCCTGTCACTAGGCCATGGTCATCCCG
ACTGTGCCCTTCAACATCACCATCAACAGCAAGCCCTTAGGACACATCTCCTTTCAGCTATTTGCAGACA
AATTTCAAAGACAGGAGAAAACCTTCACACTCTGAACAATAAAGACAAAGGATTTGGTTCCTGCTTCA
CAGAATTATTCGGAGTTTATATGCCAGGGTGATGACTTCACACCCATAATGGCATTTGGTGGCAAGTCC
ATCTACGGGGATAAATTTGATGATAAGAACTTTATTGTGAAGCATACAGGTCTTGGCATCTTGTCATGG
CAAATGCTGCACCCAAAACAAATGAGTCCAGTTTTTCATCTGCACTGCCATGGCCAAATGGTGGGATGG
CAAGCATGTGATCTTTGGCAGGGTGAAAGAGGGCATGAATATTGTGGAAGCCATGGAATGCTTTGGGTCC
AGGAATGGCAAGACAAGCAAGATCGCCATTGCCAACTGCAGACAACCTCTGATAAATTTGACTTGTGTTTT
ATCTTAACCACCAGACCTTT

In a search of public sequence databases, the NOV42a nucleic acid sequence, located on chromosome 11 has 338 of 387 bases (87%) identical to a gb:GENBANK-ID:AK026569|acc:AK026569.1 mRNA from Homo sapiens (Homo sapiens cDNA: FLJ22916 fis, clone KAT06406, highly similar to HSCYCR Human mRNA for T-cell cyclophilin).

Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV42a polypeptide (SEQ ID NO:96) encoded by SEQ ID NO:95 has 160 amino acid residues and is presented in Table 42B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV42a has no signal peptide and is likely to be localized in the microbody with a certainty of 0.6400.

5

Table 42B. Encoded NOV42a protein sequence (SEQ ID NO:96).

MVIPTVPFNITINSKPLGHISFQLFADKFPKTGENFHTLNNKDKGFGSCFHRIIPEFICQ GDDFTPHNGIGGKSIYGDKFDDKNFIVKHTGLGILSMANAAPKTNESQFFICTAMAKWWD GKHVIFGRVKEGMNIVEAMECFGSRNGKTSKIAIANCRQL
--

A search of sequence databases reveals that the NOV42a amino acid sequence has 118 of 164 amino acid residues (71%) identical to, and 135 of 164 amino acid residues (82%) similar to, the 165 amino acid residue ptnr:pir-id:CSHUA protein from human (peptidylprolyl isomerase (EC 5.2.1.8) A). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

10

NOV42a is expressed in at least Brain and Peripheral Blood. Expression information was derived from the tissue sources of the sequences that were included in the derivation of the sequence of CG57226_01. The sequence is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:AK026569|acc:AK026569.1) a closely related Homo sapiens cDNA: FLJ22916 fis, clone KAT06406, highly similar to HSCYCR Human mRNA for T-cell cyclophilin homolog in species Homo sapiens : signet-ring cell carcinoma cell_line. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

15

20

NOV42b

A disclosed NOV42b nucleic acid of 600 nucleotides (also referred to as CG57226-02) encoding a Cis/Trans Peptidyl Prolyl Isomerase -like protein is shown in Table 42C. The start and stop codons are in bold letters.

25

Table 42C. NOV42 nucleotide sequence (SEQ ID NO: 97).

CTGTGCACTGTTGGTGGGAATATAAAATGATGCAGCTGGCTTTGCAGACACTGCTGTCCC CCAACACCCCCCTGTCACTAGGCC AT GGTCATCCCGACTGTGCCCTTCAACATCACCATCA ACAGCAAGCCCTTAGGACACATCTCCTTTTCAGCTATTTGCAGACAAATTTCCAAAGACAG GAGAAAAC TTT CACACTCTGAACAATAAAGACAAAGGATTTGGTTCCTGCTTTTCACAGAA TTATTCCGGAGTTTATATGCCAGGGTGATGACTTCACACCCCATATGGCATTGGTGGCA
--

AGTCCATCTACGGGGATAAATTTGATGATAAGAACTTTATTGTGAAGCATACAGGTCTTG
GCATCTTGTCCATGGCAAATGCTGCACCCAAAACAAATGAGTCCCAGTTTTTCATCTGCA
CTGCCATGGCCAAATGGTGGGATGGCAAGCATGTGATCTTTGGCAGGGTGAAAGAGGGCA
TGAATATTGTGGAAGCCATGGAATGCTTTGGGTCCAGGAATGGCAAGACAAGCAAGATCG
CCATTGCCAACTGCAGACAACCTCTGATAAATTTGACTTGTGTTTTATCTTAACCACCAGA

In a search of public sequence databases, the NOV42b nucleic acid sequence, located on chromosome 11 has 335 of 382 bases (87%) identical to a gb:GENBANK-ID:AK026569|acc:AK026569.1 mRNA from Homo sapiens (Homo sapiens cDNA: FLJ22916
5 fis, clone KAT06406, highly similar to HSCYCR Human mRNA for T-cell cyclophilin). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV42b polypeptide (SEQ ID NO:98) encoded by SEQ ID NO:97 has 160 amino acid residues and is presented in Table 42D using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV42b has no signal peptide and is
10 likely to be localized in the microbody with a certainty of 0.6400.

Table 42D. Encoded NOV42b protein sequence (SEQ ID NO: 98).

MVIPTVPFNITINSKPLGHISFQLFADKFPKGTGENFHTLNNKDKGFGSCFHRIIPEFICQ
GDDFTPHNGIGGKSIYGDKFDDKNFIVKHTGLGILSMANAAPKTINESQFFICTAMAKWWD
GKHVIFGRVKEGMNIVEAMECFGSRNGKTSKIAIANCRQL

A search of sequence databases reveals that the NOV42b amino acid sequence has 118 of 164 amino acid residues (71%) identical to, and 135 of 164 amino acid residues (82%)
15 similar to, the 165 amino acid residue ptnr:pir-id:CSHUA protein from human (peptidylprolyl isomerase (EC 5.2.1.8) A). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV42b is expressed in at least Brain, and Peripheral Blood. This information was derived by determining the tissue sources of the sequences that were included in the invention
20 including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV42a polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 42E .

Table 42E. BLAST results for NOV42a

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 12804335 gb AAH03026.1 AAH03026 (BC003026)	protein for IMAGE:2823490) [Homo sapiens]	174	118/164 (71%)	135/164 (81%)	1e-60

gi 10863927 ref NP_066953.1 (NM_021130)	peptidylprolyl isomerase A (cyclophilin A) [Homo sapiens]	165	118/164 (71%)	135/164 (81%)	8e-60
gi 4033689 sp P04374 CYPH_BOVIN	PEPTIDYL-PROLYL CIS-TRANS ISOMERASE A (PPIASE) (ROTAMASE) (CYCLOPHILIN A) (CYCLOSPORIN A-BINDING PROTEIN)	164	118/164 (71%)	135/164 (81%)	8e-60
gi 13937981 gb AAH07104.1 AAH07104 (BC007104)	peptidylprolyl isomerase A (cyclophilin A) [Homo sapiens]	165	118/164 (71%)	135/164 (81%)	9e-60
gi 14743515 ref XP_017224.2 (XM_017224)	hypothetical protein XP_017224 [Homo sapiens]	165	118/164 (71%)	135/164 (81%)	3e-59

Table 42F lists the domain descriptions from DOMAIN analysis results against NOV 42. This indicates that the NOV42 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 42F Domain Analysis of NOV42	
gnl Pfam pfam00160 , pro_isomerase, Cyclophilin type peptidyl-prolyl cis-trans isomerase.	
CD-Length = 162 residues, 100.0% aligned	
Score = 185 bits (469), Expect = 2e-48	

The disclosed NOV42 nucleic acid of the invention encoding a Cis/Trans Peptidyl Prolyl Isomerase-like protein includes the nucleic acid whose sequence is provided in Table 42A or 42C or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 42A or 42C while still encoding a protein that maintains its Cis/Trans Peptidyl Prolyl Isomerase-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

20

In the mutant or variant nucleic acids, and their complements, up to about 13 percent of the bases may be so changed.

The disclosed NOV42 protein of the invention includes the Cis/Trans Peptidyl Prolyl Isomerase-like protein whose sequence is provided in Table 42B or 42D. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 42B or 42D while still encoding a protein that maintains its Cis/Trans Peptidyl Prolyl Isomerase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 29 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this Cis/Trans Peptidyl Prolyl Isomerase-like protein (NOV42) may function as a member of a “Cis/Trans Peptidyl Prolyl Isomerase family”. Therefore, the NOV42 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV42 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the Cis/Trans Peptidyl Prolyl Isomerase-like protein (NOV42) may be useful in gene therapy, and the Cis/Trans Peptidyl Prolyl Isomerase-like protein (NOV42) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from CNS disorders, brain disorders including epilepsy, eating disorders, schizophrenia, ADD; cancer; heart disease; inflammation and autoimmune disorders including Crohn’s disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, blood disorders; psoriasis colon cancer, leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, pancreatic disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer, or other pathologies

or conditions. The NOV42 nucleic acid encoding the Cis/Trans Peptidyl Prolyl Isomerase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV42 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV42 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV42 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV43

A disclosed NOV43 nucleic acid of 3146 nucleotides (also referred to as CG57538-01) encoding a ceruloplasmin-like protein is shown in Table 43A. The start and stop codons are in bold letters.

Table 43A. NOV43 nucleotide sequence (SEQ ID NO:99).

AATTTGAAAATGAAGGCACTTTTACCATTGACCTTCTGTTTTTTATTAGTTCTCCAGGTTGGGCAATAG
 ATAGGCACTGCTACATAGGCATTGAAGAAAGCATTGGAACATGCTAATGCTGATGAAAACCTTTCTCAT
 GATTGACACTTGCAGGACACATATGCCATTATTTCTACAAGGAGGTCAAGCGAGGAAGAGCTTTGTTTTT
 AAAAAGGCTTTGTATTTCAATATACTGATAATACATTTCAAAGGATCATTGAAAAACCATCTGGTTGG
 GATTTT TAGGTCCAATGATTAAAGCAGAGACTGGAGACTTCATTTATGTACATGTAAAAATAATGCTTC
 AAGAGCTTATAGTTATCATCCTCATGGGCTCACCTACTCCAAGAAAAATGAAGGTGCTATCTATCCTGAT
 AATACGACAGGCCTGCAAAAGGAAGATGAATATCTGGAGCCAGGGAACAATATACCTACAAGTGGTATG
 TAGAAGAACATCAGGGACCTGGCCCAATGACAGTAATTGTGTGACAAGAATTTACATTCCCATATAGA
 CACTGCAAGAGATGTAGCTTCGGGACTTATTGGACCAATACTGACTTGTAAGAGAGGTACACTGAATGGA
 GACACTGAAAAAGATATTGACAGGTCTTCTTTCTGATGTTTTCTACAACCTGATGAAAGCAGAAGCTGGT
 ATAGTGATGAAAATATTTCGTGCATTACTGAATCTGGCAAGATTAATACTAGTGATCCCCGTTTGGAGGA
 GAGCATGAGCATGCAAGCAATAAATGGATACATCTATGGAAATCTGCCCAATCTCACCATGTGTGCTGAA
 GATAGGGTCCAGTGGTATTTTGTGGCATGGGTGGCGTGGCTGACATACACCCCGTCTACCTCCGCGGAC
 AAATCTGATCTCTCGGAATCACAGAAAGGACACCATTATGCTCTTCCCCTCCTCACTGGAAGATGCCTT
 CATGGTGGCCAAGGCCCTGGAGTGTGGATGCTGGGATGCCAGATGCAGGCATTTTTCAAAGTAAGTAAT
 TGCCAGAAACCTTCAACAGAAGCCTTTGTTACTGGGACACATGTTATACATTACTATATTGCTGCTAAAG
 AAATCTTTGGAACATGCTCCATCTGGTATAGATTTCTTCACTAAAAAAATTTAACAGCAGCTGGAAG
 TAAATCCAGTTATTTTTTGAACGAAGTCCAACCAGAATTGGAGGAACTAACAAAAACTGATTTACCGT
 GAATACACAGATGCTTCTTCCAAACACAGAAGCAAGAGAAGAACACCTTGAATCCTAGGCCCGTTA
 TTAAGGCAGAGGTGAGACAGACCATCAAAATCACTTTCTATAACAATGCTTCCCTGCCACTCAGCATTC
 GCCTCCTGGATGCATTACAACAAGAGCTTGTGGCAGAGTTATTACTTTAGTTCTTATCAACTGTCAAC
 CAAAGAGAAAGATCTGTTCTCCACCTCTTACATGTAAGTCTGGCACAACATTTGTTCTATACATGGG
 AAGTTCCAAAAGATGTGGGTCCACCTCCACAGATCCCAACTGCTTGACCTGGTTCTATTACTCTTCAGT
 AAATGGGAAAAAGACATCAACAGTGGCCTTCTGGGGCCTCTCCTTATATGTAGAAATGGAAGTCTTGA
 GACGATGGCAACAGAAAGGAGTAGACAAAGAGTTTACCTACTTGCCACAATATTTGATGAAAATGAAA
 GTAATCTCTTGGATGAAAATATCAGAACATTTATCAGAGCCTGAAAACATAGATAAAGAGGATACAGA
 CTGCCAAGCCTCAAATAAGATGTACGCCATAAATGGATACATGTATGGAATCTGCCTGGATTGGACACG
 TGCTTAGGAGACAACGTTTTGTGGCAGTTTTTAGTGTAGGATCAGTGGAAAGATTTACACGGGATATATT
 TTTCAGGAAATACCTTCACTTCTTAGGAGCAAGAAGGGACACAATACCTATGTTTCTTATACTTCTCA
 GACGCTTTTGATGACACCTGATTCTATAGGTACTTTGATTGGTTTGCATGACAATAAAGCACAATCTA

GGAGGCATGAAACATAAATATCACGTGAGGCAATGTGGGAAGCCAAACCCTGATCAAACACAATACCAGG
AGGAGAAAATAATTATTACCATTCAGCCGAGGAAATGGAATGGGATATTCTCCTAGTAGAAAGTGGGA
GAATGAACCTCCACCACCTACGAAGAGAGAGCCAAACGAGCATGTATGTGGACAGAAGTGGAACTTCTT
GGGTCCAAATACAAGAAAGTCTTATATCGTCAATATGATGATAACACGTCACAAATCAAACAAAAGGAA
TGAGGGTGAAAAACATCTCGATACTAGGTCCATTAATATTGCTCAACCCTGGTCAAATAATTCAAATTAT
CTTTAAAAATAAAGCCGCAAGACCGTATTCTATTTCATGCTCATGGAGTGAAAACAAATAATCCACTGTT
GTTCCAACCTCAGCCAGGTGAGATTCAAATATATACTTGGCAGATACCTGATAGAAGTGGTCTACCTCAC
TGGACTTTGAATGCATACCTTGGTTTTACTATTCAACTGTATCTGTGGCTAAGGACCTTCACAGTGGACT
GGTAGGCCCTCTCTCTGTATGCCGCAAGACATCAACCCCAACATAGTTCACCGTGTCTCCACTTCATG
ATATTTGATGAGAATGAATCCTGGTACTTCGAAGACAGTATCAACACCTATGCTTCAAACCAACAAAG
TGGACAAGGAAATGATAATTTCAACTCAGCAACCAATGCACGCAATTAACGGAAGACTGTTTGGAAA
TAACCAAGGTATAACATTCCATGTTGGGGATGTAGTGAATTGGTATCTGATTGGCATAGGGAATGAAGCT
GGAGTGTATCAATCTGATGTTTATGACCTTCTCTCTGGGGTCTATCGAACTGTAAAAATGTATCGAAGAG
ATGTTTGAACCTGGTTATTTTATGTCATGTTTTTGAGCACATTGGTGTCTGGAATGGAAGCACTTACAC
TGTACTTGAAGAAAAGTAAGATCCATTGGCTAAATTAATTAGAAGTGATATTTAAACAAATGCA

In a search of public sequence databases, the NOV43 nucleic acid sequence, located on chromosome 3 has 1113 of 1697 bases (65%) identical to a gb:GENBANK-

ID:HUMCERP|acc:M13699.1 mRNA from Homo sapiens (Human ceruloplasmin

5 (ceruloplasmin) mRNA, complete cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV43 polypeptide (SEQ ID NO:100) encoded by SEQ ID NO:99 has 1036 amino acid residues and is presented in Table 43B using the one-letter amino acid code.

Signal P, Psort and/or Hydropathy results predict that NOV43 has a signal peptide and is
10 likely to be localized extracellularly with a certainty of 0.4085. The most likely cleavage site for a NOV43 peptide is between amino acids 19 and 20.

Table 43B. Encoded NOV43 protein sequence (SEQ ID NO:100).

MKALLPLTFLFFISSPGWAIDRHCIYIGIEESIWNANADENFLMIDTCRTHMPLFLQGGQ
ARKSFVFKKALYFQYTDNTFQRIIEKPSWLGLFGPMIKAETGDFIYVHVKNNASRAYSYH
PHGLTYSKENEGAIYPDNTTGLQKEDEYLEPGKQYTYKWYVEEHQGPNDNSNCVTRIYH
SHIDTARDVASGLIGPILTCRGTNLNGDTEKIDIRSSFLMFSTTDESRSWYSDENIRAF
ESGKINTSDPRFEESMSMQAINGYIYGNLPLNLTMCAEDRVQWYFVGMGGVADIHPVYLRG
QTLISRNRKDTIMLFSSLEDAFMVAKAPGVWMLGCMQAFKVSNCQKPSTEAFVTGT
HVIHYIIAAKEILWNYAPSGIDFFTCKNLTAAGSKSQLFFERSPTRIGGTNKKLIYREYT
DASFQTQKAREEHLGILGPVIAEVRQTIKITFYNNASLPLSIQPPGLHYNKSLSWQSYF
SSYSTVTQRERSVPPSSSHVSPGTTTFVYTWEVPKDVGPSTDPNCLTWFFYSSVNGKKDI
NSGLLGPLLCRNGSLGDDGKQKGVDFEYLLATIFDENESNLLDENIRTFITEPENIDK
EDTDCQASNKMYAINGYMYGNLPLGLDTCGLDNVLWHVFSVGSVEDLHGIYFSGNTFTSLG
ARRDTIPMFYTSQTLLMTPDSIGTFDLVCMTIKHNLGGMKHKYHVRQCCKPNPDQTYQ
EEKIIITIAAEMEWDYSPSRKWENELHHLRRESQTSMYVDRSGTLLGSKYKKVLYRQYD
DNSTSIKQKGMRVKNISILGPLILLNPGQIIQIIIFKNKAARPYSIHAHGVKTNNSTVVP
QPGEIQIYTWQIPDRGTGPTSLDFECIPWFFYSTVSVAKDLHSGLVGPLSVCRKDINPNIV
HRVLHFMIFDENESWYFEDSINTYASKPNKVDKENDNFQLSNQMHAINGRLFGNNQGITF
HVGDVVNWYLLIGIGNEAGVYQSDVYDLPPGVYRTVKMYRRDVGTLWLFYCHVFEHIGAGME
STYTTLERKKGKIHWLN

A search of sequence databases reveals that the NOV43 amino acid sequence has 548 of 994 amino acid residues (55%) identical to, and 719 of 994 amino acid residues (72%)

15 similar to, the 1069 amino acid residue ptnr:pir-id:KUH protein from human (ceruloplasmin

(EC 1.16.3.1) precursor [validated]). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV43 is expressed in at least salivary glands. Expression information was derived from the tissue sources of the sequences that were included in the derivation of the sequence of CG57538-01. The sequence is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:HUMCERP|acc:M13699.1) a closely related Human ceruloplasmin (ceruloplasmin) mRNA, complete cds homolog in species Homo sapiens :liver, secreted into plasma.. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV43 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 43C.

Table 43C. BLAST results for NOV43					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 1070458 pir KUH U	ferroxidase (EC 1.16.3.1) precursor - human	1069	579/1068 (54%)	760/1068 (70%)	0.0
gi 4557485 ref NP_0 00087.1 (NM_000096)	ceruloplasmin (ferroxidase); Ceruloplasmin [Homo sapiens]	1065	578/1066 (54%)	758/1066 (70%)	0.0
gi 1942284 pdb 1KCW I	X-Ray Crystal Structure Of Human Ceruloplasmin At 3.0 Angstroms	1048	568/1046 (54%)	746/1046 (71%)	0.0
gi 5281319 gb AAD41 477.1 AF134814.1 (AF134814)	(AF134814) ceruloplasmin [Ovis aries]	1048	577/1054 (54%)	742/1054 (69%)	0.0
gi 6680997 ref NP_0 31778.1 (NM_007752)	gi 6680997 ref NP 031778.1 (NM_007752)	1062	560/1062 (52%)	737/1062 (68%)	0.0

Table 43D lists the domain descriptions from DOMAIN analysis results against NOV43. This indicates that the NOV43 sequence has properties similar to those of other proteins known to contain this domain.

Table 43D. Domain Analysis of NOV43

gnl|Pfam|pfam00394, Cu-oxidase, Multicopper oxidase. Many of the proteins in this family contain multiple similar copies of this plastocyanin-like domain.

CD-Length = 135 residues, 90.4% aligned

Score = 37.7 bits (86), Expect = 0.003

In Wilson disease, the basal ganglia and liver undergo changes that express themselves in neurologic manifestations and signs of cirrhosis, respectively. A disturbance in copper metabolism is somehow involved in the mechanism. Low ceruloplasmin is found in the serum.

Shokeir and Shreffler (1969) advanced the hypothesis that ceruloplasmin functions in enzymatic transfer of copper to copper-containing enzymes such as cytochrome oxidase.

Supporting the hypothesis was the finding of markedly reduced levels of activity of cytochrome oxidase in Wilson disease and moderate reductions in heterozygotes. An abnormality of ceruloplasmin seems to be involved in Wilson disease. The fact that

individuals with hereditary ceruloplasmin deficiency have profound iron accumulation in most tissues suggests that ceruloplasmin is important for normal release of cellular iron

(Mukhopadhyay et al., 1998). At least 3 variants determined by codominant alleles have been identified by starch gel electrophoresis (Shreffler et al., 1967). Human ceruloplasmin is composed of a single polypeptide chain (Takahashi et al., 1984).

The disclosed NOV43 nucleic acid of the invention encoding a ceruloplasmin-like protein includes the nucleic acid whose sequence is provided in Table 43A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 43A while still encoding a protein that maintains its ceruloplasmin-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 35 percent of the bases may be so changed.

The disclosed NOV43 protein of the invention includes the ceruloplasmin-like protein whose sequence is provided in Table 43B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 43B while still encoding a protein that maintains its ceruloplasmin-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 45 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this ceruloplasmin-like protein (NOV43) may function as a member of a “ceruloplasmin family”. Therefore, the NOV43 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV43 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the ceruloplasmin-like protein (NOV43) may be useful in gene therapy, and the ceruloplasmin-like protein (NOV43) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from Wilson disease, dementia, diabetes, retinal degeneration, neurologic degeneration, xerostomia, or other pathologies or conditions. The NOV43 nucleic acid encoding the ceruloplasmin-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV43 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV43 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV43 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in

KNIRAVNLGLNHLDSVPTTLGALKELHEVGLHDNLLNNIPVSIKLPKLLKLNKRNPPF
KPGESEIFIDSIRRLLENLYVVEEKDLCAACLRKCQNARDNLNRIKNMATTTPRKTIFPNL
ISPNSMAKDSWEDWR

A search of sequence databases reveals that the NOV44a amino acid sequence has 211 of 255 amino acid residues (82%) identical to, and 235 of 255 amino acid residues (92%) similar to, the 262 amino acid residue ptmr:TREMBLNEW-ACC:BAB29635 protein from
5 Mus musculus (Mouse) (ADULT MALE TESTIS CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY, CLONE:4921523N16, FULL INSERT SEQUENCE). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV44a is expressed in at least cervix, brain, and testis. This information was derived by determining the tissue sources of the sequences that were included in the invention
10 including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

NOV44b

A disclosed NOV44b nucleic acid of 847 nucleotides (also referred to as CG57623-01) encoding a Leucine rich repeat -like protein is shown in Table 44C. The start and stop
15 codons are in bold letters.

Table 44C. NOV44b nucleotide sequence (SEQ ID NO:103).

CTTCTAACACTCCCTTACTAAAAGAACATGGTTAAGGGTGAGAAAGGCCCAAGGGCAAG
AAGATCACCCTCAAGGTGGCCAGGAATTGCATCAAAATCACTTTTGATGGGAAAAAGCGC
CTTGACTTGAGCAAGATGGGAATTACACCTTCCCAAGTGATTCTGCGCCTTAGTGAC
ATGGACGAGCTGGACCTTAGCCGAATCTTATCAGGAAGATCCCTGACTCCATCTCCAAG
TTCCAGAACCTCCGGTGGCTGGACCTGCACAGCAACTACATAGACAAGCTGCCTGAGTCC
ATTGGCCAGATGACCAGCCTGCTCTACCTCAACGTCAGCAACAACCGGCTGACCAGCAAC
GGGCTGCCCGTGGAGCTGAAGCAACTCAAGAACATCCGCGCTGTGAACCTAGGCTTGAAC
CACCTGGACAGCGTGCCACCATGCTGGGGGCCCTGAAGGAGCTCCACGAGGTAGGGCTC
CATGACAACCTACTGAACAACATCCCGTGAGCATCTCCAAGCTCCCAAGCTGAAAAAG
CTCAACATAAAGCGGAACCCCTTTCCAAAGCCAGGTGAGTCGGAATATTCATAGACTCC
ATCAGGAGGCTGGAGAACTTGATGTGTGGAGGAGAAGGATCTGTGTGCGGCTTGCTG
AGAAAATGCCAAACGCCGGGACAACTGAATAGAATCAAGAACATGGCCACGACGACA
CCGAGAAAGACCATCTTTCCCAATCTGATCTCACCAATTCATGGCCAAGGACTCCTGG
GAAGACTGGAGG**TGACT**TGGAACCTGAGCCCTGAGGCAGAAAGGAAAGAGAGAGGGAGGG
AAGAGGG

In a search of public sequence databases, the NOV nucleic acid sequence, located on chromosome 10 has 187 of 313 bases (59%) identical to a gb:GENBANK-
20 ID:AB016816|acc:AB016816.1 mRNA from Homo sapiens (Homo sapiens MASL1 mRNA, complete cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

<u>gi</u> 12838360 <u>dbj</u> BAB24176.1 (AK005666)	Leucine Rich Repeat containing protein-data source: Pfam, source key: PF00560, evidence: ISS-putative [Mus musculus]	230	195/228 (85%)	214/228 (93%)	e-97
<u>gi</u> 10130019 <u>gb</u> AAG13461.1 AF274972.1 (AF274972)	PIDD [Homo sapiens]	910	48/146 (32%)	87/146 (58%),	2e-18
<u>gi</u> 12083587 <u>ref</u> NP073145.1 (NM_022654)	p53 protein induced, with death domain [Mus musculus]	915	47/146 (32%)	87/146 (59%)	2e-18

Leucine-rich repeats (LRRs) are relatively short motifs (22-28 residues in length) found in a variety of cytoplasmic, membrane and extracellular proteins. Although these proteins are associated with widely different functions, a common property involves protein-protein interaction. Little is known about the 3D structure of LRRs, although it is believed that they can form amphipathic structures with hydrophobic surfaces capable of interacting with membranes. In vitro studies of a synthetic LRR from *Drosophila* Toll protein have indicated that the peptides form gels by adopting beta-sheet structures that form extended filaments. These results are consistent with the idea that LRRs mediate protein-protein interactions and cellular adhesion. Other functions of LRR-containing proteins include, for example, binding to enzymes and vascular repair. The 3-D structure of ribonuclease inhibitor, a protein containing 15 LRRs, has been determined, revealing LRRs to be a new class of alpha/beta fold. LRRs form elongated non-globular structures and are often flanked by cysteine rich domains.

The disclosed NOV44 nucleic acid of the invention encoding a leucine-rich repeat-like protein includes the nucleic acid whose sequence is provided in Table 44A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 44A while still encoding a protein that maintains its leucine-rich repeat-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense

The disclosed NOV44 protein of the invention includes the leucine-rich repeat-like protein whose sequence is provided in Table 44B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table B while still encoding a protein that maintains its leucine-rich repeat-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 18 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this leucine-rich repeat-like protein (NOV44) may function as a member of a “leucine-rich repeat family”. Therefore, the NOV44 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV44 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the leucine-rich repeat-like protein (NOV44) may be useful in gene therapy, and the leucine-rich repeat-like protein (NOV44) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from fertility, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, or other pathologies or conditions. The NOV44 nucleic acid encoding the leucine-rich repeat-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

antibodies that bind immuno-specifically to the novel NOV44 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV44 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV45

NOV45 includes two Ig/fibronectin-like proteins disclosed below. The disclosed sequences have been named NOV45a and NOV45b.

NOV45a

A disclosed NOV45a nucleic acid of 4321 nucleotides (also referred to as CG57656-01) encoding a Ig/fibronectin -like protein is shown in Table 45A. The start and stop codons are in bold letters.

Table 45A. NOV45a nucleotide sequence (SEQ ID NO: 105).

CTGGTGGGGGGCGGGGTGACCTGTGACACGGACATGGGGCTGCTGGGGCAGGATCTCTTTGTACCTCC
TTTCTGTGTCCAACCTGGCCGTCCCCATCAGGCGCCACGCGCTGCGAGAGGAGCCCGAGTTTGTGACGG
CAAGAGCTGGGGAGAGCGTGGTCTTCGCGATGCGACGTGATCCACCCAGTGACGGGACAGCCCCACCCCTA
TGTCGTAGAGTGGTTCAAGTTCGGGGTCCCCATCCCTATCTTCATCAAGTTTGGCTACTACCCGCCGCAC
GTGGACCTGAGTATGACAGTAAGGTGCGGCGCCACGGCCTGCGAGAGGAGCCCGAGTTTGTGACGGCAA
GAGCTGGGGAGAGCGTGGTCTTCGCGATGCGACGTGATCCACCCAGTGACGGGACAGCCCCACCCCTATGT
CGTAGAGTGGTTCAAGTTCGGGGTCCCCATCCCTATCTTCATCAAGTTTGGCTACTACCCGCCGCACGTG
GACCCCTGAGTATGACAGTAAGGTGAGTCTTCATGATAAGGCATCTCTGCGGCTGGAACAAGTTCGCTCTG
AGGACCAGGGCTGGTATGAGTGCAAGTGCTCATGCTGGACACAGCAGTATGACACCTTCCACAATGGCAG
CTGGGTCCACCTCACCATCAACGCCCTCCACCTTTACAGAAACACCCCCCAGTACATCGAGGCCAAG
GAGGGTGGTAGTATCACCATGACCTGCACAGCTTTTGGGAACCCCAAGCCCATGTCACTGGCTCAAGG
AGGGGACGCTCCTCGGTGCTAGTGGGAAATACCAGGTGAGTGTGGTTCTAGGTAGCCTGACAGTGACATC
GGTCAGTTCGGGAGGACAGAGGTGCCTACACCTGCCGAGCGTACAGCATTACGGGGGAGGCTGTCCACACG
ACTCACCTGCTTGTCCAAGGGCCCCCTTTTCATCGTCTCCCTCCTGAGAACATCACCGTCAACATCTCCC
AGGATGCTCTGCTCACCTGCCGGGAGAGGCGTATCCGGGCAACCTCACCTACACCTGGTACTGGCAGGA
CGAGAACGTCTACTTTGAGAAGCAGCTGAAGCTGAGGGTGCGCATCCTAATCGATGGGACCTGATCATC
TTCCGGGTGAAGCCGAGGAGTACAGGGAAGTACACCTGTGTGCCAGCAACAGCCTGGGGCGCTCCCCCT
CCGCTCGGGCTACCTGACCGTGCAGTACCCAGCGGTGTCTCAACATGCCCTGTGATTTACGTGCC
CGTGGGGATCCATGGCTACATCCGCTGCCCTGTGGACGACAGAACACCGGCCACCGTGGTCAAGTGAAC
AAGGACGCGCCCTCCCCTGAGGTTGAGAAGAAGCTCGGTGGACCTGATGGAGGATGGCTCCATTGAA
TTGAGGAGGCCACAGAGGAGGCTCTTGGCACTTATACCTGTGTGCCTTACAACACTCTGGGGACCATGGG
CCAGTCTGCCCTGCGAGGCTTGTCTGAAGGACCCCCCTATTTCACGGTGTACACAGGTTGGGAGTAC
AGGCAGGAGGCCGCGCGGAGCTACTTATCCCTGTGCTGCCGAGGGGACCCCTTTCTGTCTCACTT
GGAGAAAGGTAGGGAAGCCAGCAGAAGCAAGCAGTCCCTGCCAGTGGGAGCCTGCAGTTCCGTGC
CCTGAGTAAGGAGGACACGGGGAGTGGGAATGTGTGCCACCAACGTGGTACAGAGCATCACTGCCAGC
ACCCACCTCACCGTCATCGGTACGGGACACGCCCTATGCCCGGGCAGTGTCCGGGTCCAGGTCTCCA
TGACAACCTGCCAACGTGTCTGGGAACAGGTATGATGGAGGCTACAGAGCAGACATTCTCAGTTTGGTA
CGGACCTCTGATGAAGCGGGCAGTCTTGGGCCCCATGACTGGCTGTCTTCCAGTGCCGCCAGGACCC
AGCTGGCTGCTGGTGGACACCTGGAGCCTGAGACAGCGTACCAGTTACAGCTCCTGGCCAGAAGCTGG

GAACCAGCGCCTTCAGTGAGGTGGTCACTGTGAACACTTTAGCATTCCCTATTACAACCTCCAGAACCCT
GGTGCTGGTCACCCACCGAGGTGCCTCATAGCCAATCGGACTCAGCAGGGTGTGCTCCTGTCTGGCTT
CCGCTGCCAACACAGCTTCCCATCGACCGCTACATCATGGAGTTCGGTGTGCGAGAGCGCTGGGAGT
TGCTCGACGATGGCATCCCCGGCACCGAAGGAGAGTTCTTTGCCAAGGATCTGTACAGGACACGTGGTA
TGAGTTCGGGTCTGCGCGTCAATGCAGGATCTGATCGGCGAGCCAGCAACATCGCCGGCGTCTCCAGC
ACAGACATCTTCCCGCAGCCGACCTGACCGAGGATGGGCTGGCGCGGCTGTGCTGGCGGGAATCGTAG
CTACCATCTGCTTCTTGGCAGCTGCCATCCTGTTTCCAGCACCTGGCTGCCTGCTTTGTCAACAAGCAGCG
CAAGCGTAAGCTCAAGCGCAAAAAGACCTCCACTCTCCATCACCCACTGCAGGAAGAGCCTGGAGTCT
CCCTTGCTCTCTGGCAAGGTGAGCCCCGAGAGCATCCGCACGCTCCGAGCGCCGTGAGAATCTCCGACG
ACCAGGGCCAGCCCGCGCCAGAGGATGCTGAGCCCCACCCGTGAGAAGGAGCTGTGCTGTACAAGAA
GACCAAGCGGGCCATCAGCAGCAAGAAGTACAGCGTGGCCAAGGCAGAGGCCGAGGCAGAGGCCACACG
CCCATCGAGCTCATCAGCAGAGGCCCTGACGGCCGCTTCGTGATGGACCTGTGAGATGGAGCCCTCGC
TGAGAGCAGGCGCATCGAGGGCTTCCCTTCGCCGAGGAGACGGACATGATACCCGAGTTCGCCAGTC
GGACGAGGAGAACGAGGACCACTGGTGCCACATCTGTGGCCGCTGAAGTCCCAGCTCACCCCTCTG
TCATCCAGCCAGGAGTCTTACCTGCCACCACAGCATAACAGCCCTCGGTTCCAGCCCCGCGGGTGGAGG
GCCCCGGTGGCTGGAAGGTGCGCTTACGGCCACAGGCCAGGCCGCGCCCTGCCCCCGGCCCTTCCA
CCATGGCCAGTATTATGGGTACCTCAGCAGCAGCAGCCCTGGGGAGGTGGAGCCGCCCGCTTCTACGTG
CCAGAAGTGGGCGAGCCCCCTGAGCTCCGTCTATGTCGTCCCCGCCCTGCCACCGAGGGGCCCTTTGGCC
ACCCACCATCCCCGAGGAGAATGGAGAGAATGCATCCAACAGCAGCTGCCCTTGACTCAGACACCTAC
AGGAGGCGCTCCCCCTGAGCCCTGGGGCCGCGCCAGAATTCCTTCGGGGGGCTGGAGACCCAGCGATG
ATGTTCCCCCACCAGCTGCCACCTGTGATGTGCCCGAGAGTCTGCAGCCCAAGGCCGCGCTCCCCGAG
GACTGCCCCCCACCTCCCTGCAGGTGCCCGCGGCTTACCCGGGCATCCTGTCTCTGAGGCACCGAAGGG
TTGGGCAGGCAAGTCGCCCGGCGAGGGGCCCTGTCCAGCGCCCCCGCGCCAAGTGGCAGGACAGACCT
ATGCAACCTCTGGTAAGCCAAGGGCAGCTGCGACATACAAGCCAAGGCATGGGCATACCTGTGCTGCTT
ACCCGAGCCGGCTGAGCCGGGGCGCACGGCGGCCAGCACATTTGGCTGGACACCCGGTGGTATGA
GCCCCAGCCCCGGCCCCGCTAGCCCTCGGCAGGCCAGGCGCGCCGAGCCAGTTTACATCAAGTGGTG
CTACAGCCCTCCCGGCTCTACCTCTGACCCAAAGCCCCCTCAGCTCCCGACCGGCTCCCCCTGAGCTCG
CCGCCCCGTGCCCGGCTCGCCCGGCTCCTGCAGCAGGCAGAGATGTGAGATCACCTGCAGCCGCC
GGCTGCAGTCAGCTTTTCTCGAAAGTCTACGCCGTCCACAGGCTCCCCCTCCAGAGCAGCCGAGTGGG
AGTCCAGCTACCGGCCCGCATGGGCTTACCACCTCTGGCCACCGGCTACCCTTCCCTCCACCCGGCC
CCGCCCTGTGGGCTGGGGACAGCTTGGACGTGTTTGGACAGACGCTTCCCCCTCGAAGGACGGGGGA
GGAATTGCTCCGACCGGAGACCCACCCACGTTACCTACTTCAGGGAAGCTGCGGAGAGACAGACCA
GCTCCCGCAGCAGCCGCTGAGAGAGCACTCTTAACTGTAGCAGCTG

In a search of public sequence databases, the NOV45a nucleic acid sequence, located on chromosome 11 has 2296 of 2298 bases (99%) identical to a gb:GENBANK-
ID:AB028953|acc:AB028953.1 mRNA from Homo sapiens (Homo sapiens mRNA for
5 KIAA1030 protein, partial cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV45a polypeptide (SEQ ID NO: 106) encoded by SEQ ID NO:105 has 1328 amino acid residues and is presented in Table 45B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV45a has a signal peptide and
10 is likely to be localized at the plasma membrane with a certainty of 0.4600. The most likely cleavage site for a NOV45a peptide is between amino acids 19 and 20.

Table 45B. Encoded NOV45a protein sequence (SEQ ID NO: 106).

MGLLGQDLFVTSFLCPTWSPSPGAHGLREEPEFVTARAGESVVLRCVDVIHPVTGQPPPYV
VEWFKFGVPIPIFIKFGYPPHVDPEYAGKVGAHGLREEPEFVTARAGESVVLRCVDIHP
VTGQPPPYVVEWFKFGVPIPIFIKFGYPPHVDPEYAGKVS LHDKASLRLEQVRSEDQGW
YECKVLM LDQQYDTFHNGSWVHLTINAPPTFTETPPQYIEAKEGGSITMTCTAFGNPKPI
VTWLKEGTL LGASGKYQVSVVLGSLTVTSVSREDRGAYTCRAYSIQGEAVHTHLLVQGP
PFI VSPPENITVNISQDALLTCRAEAYPGNLT YTWYQDENVYFQNDLKLVRILIDGTL
IIFRVKPEDSGKYTCVPSNSLGRSPSASAYLTVQYPARVLNMPPVIYVPVGIHGYIRCPV
DAEPPTATVVKWNKDGRPLQVEKNLGTLMEDGSIRIEEATEEALGTYTCVPYNTLGTMGQ
SAPARLV LKDPYFTVLPGWEYRQEAGRELLIPCAAAGDPFPVITWRKVGKPSRSKHSAL
PSGSLQFRALSKEDHGEWECVATNVVTSITASTHLTVIGTGTSPHAPGSVRVQVSMTTAN

VSWEPGYDGGYEQTFSVWYGPLMKRAQFGPHDWLSLPVPPGPSWLLVDTLEPETAYQFSV
 LAQKLGTSAFSEVVTVNTLAFPIITTEPLVLVTPPRCLIANRTQQGVLLSWLPPANHSFP
 IDRYIMEFRVAERWELLDDGIPGTEGEFFAKDLSQDTWYEFRLAVMQDLIGEPSNIAGV
 SSTDIFPQPDLTEDGLARPVLGIVATICFLAAAILFSTLAACFVNKQRKRKLKRKKDPP
 LSITHCRKSLESPLSSGKVSPEIIRTLRAPSESSDDQGPAAKRMLSPTREKELSLYKKT
 KRAISSKKYSVAKAEAEAEATTPIELISRGPDGRFVMDPVEMEPSLKSRRIEGFPPAEET
 DMYPEFRQSDEENEDPLVPTSVAALKSQLTPLSSSQESYLPPPAYSPPRQPRGLEGPGL
 EGRLQATGQARPPAPRPFHHGQYYGYLSSSSPGEVEPPPFYVPEVGSPLSSVMSSPPLPT
 EGPFGHPTIPEENGENASNSTLPLTQTPTGGRSPEPWGRPEFPFGGLETPAMMFPHQLPP
 CDVPESLQPKAGLPRGLPPTSLOVPAAYPGILSLEAPKGWAGKSPGRGPVPAPPAAKWQD
 RPMQPLVSQGLRHTSQGMGIPVLPYPEPAEPGAHGGPSTFGLDTRWYEPQPRPRPSPRQ
 ARRAEPSLHQVVLQPSRLSPLTQSPLSSRTGSPELAARARPRPGLLQQAEMSEITLQPPA
 AVSFSRKSTPSTGSPSQSSRSGSPSYRPAMGFTTLATGYPSPPPAPAGPGDSLDFVFGQ
 TPSPRRTGEELLRPETPPPTLPTSGKLRRDRPAPATSPPERALSKL

A search of sequence databases reveals that the NOV45a amino acid sequence has 761 of 763 amino acid residues (99%) identical to, and 761 of 763 amino acid residues (99%) similar to, the 763 amino acid residue ptnr:SPTREMBL-ACC:Q9UPX0 protein from Homo sapiens (Human) (KIAA1030 PROTEIN). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV45a is expressed in at least brain, cerebral medulla/cerebral white matter, prostate, thalamus, placenta. Expression information was derived from the tissue sources of the sequences that were included in the derivation of the sequence of CG57656-01. The sequence is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:AB028953|acc:AB028953.1) a closely related Homo sapiens mRNA for KIAA1030 protein, partial cds homolog in species Homo sapiens :liver. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

NOV45b

A disclosed NOV45b nucleic acid of 7097 nucleotides (also referred to as CG57656-02) encoding a Ig/fibronectin -like protein is shown in Table 45C. The start and stop codons are in bold letters.

Table 45C. NOV45b nucleotide sequence (SEQ ID NO: 107).
AT GGGGCTGCTGGGGCAGGATCTCTTTGTCACCTCCTTTCTGTGTCCAACCTGGCCGTCC CCATCAGGCGCCACGGCCTGCGAGAGGAACCCGAGTTTGTGACGGCAAGAGCTGGGGAG AGCGTGGTCTCGGATGCGACGTGATCCACCCAGTGACGGGACAGCCCCACCCTATGTC GTAGAGTGGTTCAAGTTCGGGGTCCCCATCCCTATCTTCATCAAGTTTGGCTACTACCCC CCACACGTGGACCTGAGTATGCAGGCCGGCCAGTCTTCATGATAAGGCATCTCTGCGG CTGGAACAAGTTCGCTCTGAGGACCTGGGCTGGTATGAGTGCAAAGTGCTCATGCTGGAC CAGCAGTATGACACCTCCACAATGGCAGCTGGGTCCACCTCACCATCAACGCCCTCCC

ACCTTTACAGAAACACCCCCCGGTACATCGAGGCCAAGGAGGGTGGTAGTATCACCATG
ACCTGCACAGCTTTTGGGAACCCCAAGCCCATTTGTCACCTGGCTCAAGGAGGGGACGCTC
CTCGGTGCTAGTGGGAAATACCAGGTGAGTGACGGCAGCCTGACAGTGACATCGGTACAGT
CGGGAGGACAGAGGTGCCTACACCTGCCGAGCGTACAGCATTCAGGGGGAGGCTGTCCAC
ACGACTCACCTGCTTGTCCCAGGGCCCCCTTTTCATCGTCTCCCCTCCTGAGAACATCACC
GTCAACATCTCCCAGGATGCTCTGCTCACCTGCCGGGCGAGGGCGTATCCGGGCAACCTC
ACCTACACCTGGTACTGGCAGGACGAGAAGCTCTACTTTTCAAGACACCTGAAGCTGAGG
GTGCGCATCTAATCGATGGGACCTGATCATCTTCCGGGTGAAGCCGGAGGACTCGGGG
AAGTACACCTGTGTGCCCAGCAACAGCCTGGGGCGCTCCCCCTCCGCTCGGCGTACCTG
ACCGTGCAGTACCCAGCGCGTGTCTCAACATGCCCCCTGTGATTTACGTGCCCCGTGGGG
ATCCATGGCTACATCCGCTGCCCTGTGGACGCAAGACCACCGGCCACCGTGGTCAAGTGG
AACAAGGACGGCCGTCCCCCTGCAGGTTGAGAAGAACCAGCGGTTGGACCCTGATGGAGGAT
GGCTCCATTCGAATTGAGGAGGCCACAGAGGAGGCTCTTGGCACTTATACCTGTGTGCCT
TACAACACTCTGGGGACCATGGGCCAGTCTGCCCCCTGCGAGGCTTGTCTGAAGGACCCC
CCCTATTTACGGTGCTACCAGGCTGGGAGTACAGGCAGGAGGCCGGCCGGGAGCTACTT
ATCCCCGTGTGCTGCCGAGGGGACCCCTTTCCTGTCTACTTGGAGCAAGGTAGGGAAG
CCCAGCAGAAGCAAGCACAGTGCCCTGCCAGTGGGAGCCTGCAGTTCCGTGCCCTGAGT
AAGGAGGACCACGGGGAGTGGGAATGTGTGCCACCAACGTGGTCACGAGCATCACTGCC
AGCACCCACCTCACCGTCATCGGCACAGCCCCCATGCCCGGGCAGTGTCCGGGTCCAG
GTCTCCATGACAACTGCCAACGTGTCTTGGGAACCAGGTGACGGGCTACGATGGGGCTAT
GATGGAGGCTACGAGCAGACATTCTCAGTTTGGATGAAGCGGGCACAGTTTGGGCCCCAT
GACTGGCTGTCTTGGCAGTGCCGCCAGGACCCAGCTGGCTGTGGTGGACACCCCTGGAG
CCTGAGACAGCGTACCAGTTCAGCGTCTTGGCCAGAACAGCTGGGAACCAGCGCCTTC
AGTGAGGTGGTCACTGTGATCACTTTAGCATTTCCCTATTACAACCTCCAGAACCCTGGTG
CTGGTCACCCACCGAGGTGCCTCATAGCCAATCGGACTCAGCAGGGTGTGCTCCTGTCC
TGGCTTCCGCTGCCAACACAGCTTTCCCATCGACCGCTACATCATGGAGTTCCGTGTCT
GCAGAGCGCTGGGAGTTGCTCGACGATGGCATCCCCGGCACCGAAGGAGAGTTCTTTGCC
AAGGATCTGTACAGGACACGTGGTATGAGTTCCGGGTTCTGGCCGTATGCAGGATCTG
ATCGGCGAGCCAGCAACATCGCCGGCGTCTCCAGCACAGACATCTTCCCGCAGCCGGAC
CTGACCAGGATGGGCTGGCGCGGCCTGTGCTGGCGGGAATCGTAGCTACCATCTGCTTC
TTGGCAGCTGCCATCTGTTCAGCACCTGGCTGCCCTGCTTTGTCAACAAGCAGCGCAAG
CGTAAGCTCAAGCGCAAAAAAGACCCCTCCACTCTCCATCACCCACTGCAGGAAGAGCCTG
GAGTCTCCCTTGTCTCTGGCAAGGTGAGCCCCGAGAGCATCCGCACGCTCCGAGCGCCG
TCAGAATCCTCCGACGACAGGGCCAGCCCGCGGCCAAGAGGATGCTGAGCCCCACCCGT
GAGAAGGAGCTGTGCTGTACAAGAAGACCAAGCGGGCCATCAGCAGCAAGAAGTACAGC
GTGGCCAAAGCAGAGGCCGAGGCAGAGGCCACCACGCCCATCGAGCTCATCAGCAGAGGC
CTGACGGCCGCTTCGTGATGGACCTGTGAGATGGAGCCCTCGCTGAAGAGCAGGCGC
ATCGAGGGCTTCCCCTTCGCCGAGGAGACGGACATGTACCCGAGTTCCGCCAGTCCGAC
GAGGAGAACGAGGACCCACTGGTGCCACATCTGTGGCCGCCCTGAAGTCCCAGTCAAC
CCTCTGTCTCCAGCCAGGAGTCTACCTGCCACCACCAGCATACAGCCCTCGGTTCCAG
CCCCGCGGGCTGGAGGGCCCCGGTGGCCTGGAAGGTGCGCTTCAGGCCACAGGCCAGGCC
CGGCCCCCTGCCCCCCGGCCCTTCCACCATGGCCAGTATTATGGGTACCTCAGCAGCAGC
AGCCCTGGGGAGGTGGAGCCGCCCGCTTCTACGTGCCAGAAGTGGGCAGCCCCCTGAGC
TCCGTCTGTGCTCCCCGCCCCCTGCCACCAGGGGGCCCTTTGGCCACCCACCATCCCC
GAGGAGAATGGAGAGAATGCATCCAACAGCACGCTGCCCTTGACTCAGACACCTACAGGA
GGGCGCTCCCCTGAGCCCTGGGGCCGGCCAGAATTCCCCTTCGGGGGGCTGGAGACCCCA
GCGATGATGTTCCCCCACCAGCTGCCACCCTGTGATGTGCCCGAGAGTCTGCAGCCCAAG
GCCGGCTTCCCCGAGGACTGCCCCCACCTCCCTGCAGGTGCCCGCGGCCTACCCGGGC
ATCCTGTCTCTGGAGGCACCGAAGGTTGGGCAGGCAAGTCGCCCGGCAGGGGCCCTGTCT
CCAGCGCCCCCGCGCCAAGTGGCAGGACAGACCTATGCAACCTCTGGTAAGCCAAGGG
CAGCTGCGACATACAAGCCAAGGCATGGGCATACCTGTGCTGCCTTACCCCGAGCCGGCT
GAGCCGGGGCGCACGGCGGCCCCAGCACATTTGGCCTGGACACCCGGTGGTATGAGCCC
CAGCCCCGGCCCCGGCCTAGCCCTCGGCAGGCCAGGCGCGCCGAGCCAGTTTACATCAA
GTGGTGCTACAGCCCTCCCGGCTCTACCTCTGACCCAAAGCCCCCTCAGTCCCGCACC
GGTCCCCCTGAGCTCGCCGCCCGTGGCCGGCCTCGCCCGGCCTCCTGCAGCAGGAGAG
ATGTCAGAGATCACCCCTGCAGCCGCCGGCTGCAGTCAGCTTTTCTCGAAAGTCTACGCCG
TCCACAGGCTCCCCCTCCAGAGCAGCCGAGTGGGAGTCCAGCTACCGGCCCGCCATG
GGCTTACCACCTCTGGCCACCGGCTACCCCTTCCCCTCCACCCGGCCCCCGCCCTGCTGGG
CCTGGGGACAGCTTGGACGTGTTTGGACAGACGCCTTCCCCTCGAAGGACGGGGGAGGAA
TTGCTCCGACCGGAGACCCACCCACCGTACCTACTTCAGGGAAGCTGCAGAGAGAC
AGACCAGCTCCCGCGACAGCCCGCCTGAGAGAGCACTCTCTAACTGTAGCAGCTGGTA
TTCCAGCTATCTGGGCAGTGTGTGAGACAAGCCTCTCCTCAGCTCAATAGGTAAGGGA

GCTCCTCGGCTGGGCGGGCGGGGCGGGCAGGCGGACGGGGCTTCGGCCGGGCCATTGCT
 TCCTGGACAGGGGATCCAAACCATGTCCCCTACCGCCCCGTGGGGTGGCCGCTGCCGCT
 CCTCGATCCCCACGGCTTCTGGGTTCACATCGAGCCACGCTCGGCACCGCCTAGCTG
 CAGTTCTGCCCCCACCCTCGTCCCATGTCCGGCCCCCTCTAGAGCCCCCTGCGGCTC
 TCCTCACTCCTCGCCAGCCTCGCCAGCCTGCTGCCAGTACACCCAGGGCCTCCACAGAA
 GCCCTGGGGCCCTGCATGCACCCCGAAGGGGCCAGAACACCCGAGATCTGCTTTGCATC
 TCTGCACCCCTGGGGACCTCTCTGGGGCCCCCTCTGGTATCCAGGAAGAGCCCCCGCAC
 CCTATTCTCCAGCAGCCCCAGGCATAACATCCCTTCTCCTGGGGTGGCTGGCCCCCTCAC
 CAGCTTCTTGAGATCTTCGTGGTATTAGGCTCTCTCGAGGAGTACAAGTTGGGATGAGC
 CCCACTTCTTCTTGACAGTGGGGTCTCTGCAGGGTCAGGTCCCTGCAGTCTGTGGGG
 CTTTGTAGCGAAATGTCATCACTCCCCTGTGCTCCTTGCCCTTCTGCAGCCTGACCTTGGT
 AGTCACAGGACTGAAATGTGACCTAGCCCTTGGGTCCACATTGCTTTCAAGTCCCTGTT
 GTAGCCTTGTCGTCTCCTCGGTGATGCTGTAGGGAAAGGGTCTTAGGAGGACCTCCCC
 GAGGGGAGGGGGCAGGCTTCCCCTGGGCAGACAGGCATTGAGCTGGCAGGAAGTGAAACC
 CCCAGGGACACAGCACTGTGTCCCCTCCCCCTGCCCCGAGTCAGCCTCTCCTTGAGTGTCT
 CTTCAGGTGGAAGCTAAAGGAACCTGGGCATTAGGGAGCCTCTGAGTCCCTCACAGAG
 ACTCAGGGCCACCCGAAGCTCGCTCCCTCTCACAGTAGCCTGACAAGGTGCTGCCGCTCG
 CCACGGACCTCTGCCCTGTGCCTGGGCACACACAGGCATCGGGCACCTGCATGGGAGAC
 GGCGAGCCTCCCGTCCAGGTGCTTCTGGCTTTCCAGGCGAGAAGGAGACAGGTGCCTTCC
 CCCCTAGAGATGTCAAGGGAGGTTCACTTTCTTAACCAGGGCTATAAATCTCATTCATT
 CCTAAGAGTGGCCTCCATAAAGAGGACTGCCCTGACATTTCTTTACCATCTGTAGCTATA
 AAATTGTCAGCAGCCCAGAGGCCCTGGAATTGTGGGTGAGGCTGTCTGGCATAGGGTGAC
 TGTGAGGGCTGTGAGGCAGCGTTTATGAGCTCACCTGCTCTGGGGCCCTCCTGCGCCAG
 AGACCACTGGTCTTCCACCTTCCCAGCTCCCTCCCTGTCCAGCTGTGACTCCTCCAC
 AGCCCCGGGAGGTGTGGGAGCCCTAGACTAGCTCTCACCCAGCTCCTGGAAATTCCTA
 GACTCTGCTCTCATGTGTTATTCTCACCCTTCACTCAACTCCATTGACCCTCCCCTTT
 CCCAGTGTCCCCACTGTGCCAGATCCAAGAGAAGCCCGCTCTCCTTTCACTGTGTGGCA
 AACCAGAAACCAGGGGCAGCGATGAGGGACCATCGTTTCTCTCCCGGGCTGGCCAGCA
 TCCCCAGCCTAGGAGAAGGAGACCTCCCCATCCTGAGTCAGCCCCCTCGGTGCTGGCCTC
 TCCTGCCTGCTGGGAGCCTCCCCGGTACGGCTGCTGGGTCTGGGGAGATGCAGGCTCTG
 TTCAGATGCTGTCTCCATGCTGCACCTTTGCATGTGTGCCCTCTTGGTCTGTGTTGGAGA
 AGTTGTGACAGCTCTTGTCCACGTGCGTTCTCACTGTTTCTTTTCCCTTCAGTTCTCTC
 CTGTTCCCTTTGGTCCATATGTAGTTGTTGCCCTGCCTCTCCTTTTCTCTCCTCTTTT
 TTCTCCTCTTCCCACTCCCTTCTCAACATGGGGAGTTGTATTGCGTGACCTGTCTTT
 CTGTGGACTGTGGGGAGGCCGAGCCAATTCTTAGCTCCTGTCTGAGCCTAGTGATGCCCT
 GTCCTCACCTCTGCTTCCACTCCCTTTCCATTACAGGCTGACTGTACCCGACGGGAGGCA
 GGTGGGTCTGGAGCATGGAGTGGCCCCCTGTCTGGGTGAGCTGTAGGAAGGGGGCCCTTT
 GTAAACAAGCTTCTGCCAGGTTCTGAGGCAGAGCACTGGTCTGAGTCCGGCCCCACAGAA
 GGGTTCATCTTGTGGTCAGCGGGGTGGGCTCAGACCTGCAGACTCCCGCAGTGGGCCCCA
 CCCTTTCCCTTGCTATGGCCCGTCTATGAGTGGCTTTCCCTGCCTCCATCGGTATATAA
 TCAGGCAGGTGGTGTGTTGCTCTTAGGTGATTAGAAGAGGAGGAAAATGCCCATCAAGCCA
 CTGCTTCAGCCTGTTGGCCAGGGAACACCTGAACAGGGTCAAAATGGTCAGCTCCTGGGC
 GCTGAGCTAAGGAAATATGCAGGGGGTTTTGACTTTCCCTGTCTGAGGCTTGCCCTCCAC
 TTTGACCTGGCTGCATCTTTCTCAGCTCTTCTGGGTAAGCTGAGAGCTAAGCAAATAAG
 ACCTCCTGGGCTACCTGGTCATGATTTGGGTTTTGTATGGATTTTTTGAAAAGAGAGAAA
 AAGAGTGTGGGCTGGGTGCCGTGGCCAGAGGGCAGATCTCTGAGTCCAGGAGTTCAAG
 GCCAGCTGGGCAACATAGTGAAGCCCCATCTCTGCAAAAAAATACAAAAAATAGCTGG
 GCATGGTGGCACATGCCTATAGTCCCAGCTCCTTGGGGGGCTGAGGTGGGAAGATCATCT
 GAGCTGGGGAGGTGAGGCAGCAGTGAGCCGAGATTGCGCCACTTCACTCCAGCCTGGGT
 GACAGAGATCCTGTCTC

In a search of public sequence databases, the NOV45b nucleic acid sequence, located
 on chromosome 11 has 5312 of 5319 bases (99%) identical to a gb:GENBANK-
 ID:AB028953|acc:AB028953.1 mRNA from Homo sapiens (Homo sapiens mRNA for
 KIAA1030 protein, partial cds). Public nucleotide databases include all GenBank databases
 and the GeneSeq patent database.

The disclosed NOV45b polypeptide (SEQ ID NO: 108) encoded by SEQ ID NO: 107 has 1356 amino acid residues and is presented in Table 45D using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV 45b has a signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.4600. The most likely cleavage site for a NOV45b peptide is between amino acids 19 and 20 .

Table 45D. Encoded NOV45b protein sequence (SEQ ID NO: 108).

```
MGLLGQDLFVTSFLCPTWSPSPGAHGLREEPEFVTARAGESVVLRCDDVIHPVTGQPPPVV
VEWFKFGVPIPIFIKFGYPPHVDPEYAGRASLHDKASLRLEQVRSEDLGWYECKVLMLD
QQYDTFHNGSWVHLTINAPPTFTETPPRYIEAKEGGSITMTCTAFGNPKPIVTLKEGTL
LGASGKYQVSDGSLTIVTSVSREDRGAYTCRAYSIQGEAVHTTHLLVPGPPFIVSPPENIT
VNISQDALLTCRAEAYPGNLTYTWYQDENVYFQNDLKLVRILIDGTLIIFRVKPEDSG
KYTCVPSNSLGRSPSASAYLTVQYPARVLNMPPVIYVPVGIHGYIRCPVDARPPATVVKW
NKDGRPLQVEKNRGWTLMEDGSIRIEEATEEALGTYTCVPYNTLGTMGQSAPARLVLDKP
PYFTVLPGWEYRQEAGRELLIPCAAAGDPFPVITWSKVGPSPRSKHSALPSGSLQFRALS
KEDHGEWECVATNVVTSITASTHLTVIGTSPHAPGSRVRVQVSMTTANVSWEPGDGLRWGY
DGGYEQTFSVWMKRAQFGPHDWLSLPVPPGPSWLLVDTLEPETAYQFSVLAQNKLGTSAF
SEVVTVITLAFPIITTEPLVLVTPPRCLIANRTQQGVLLSWLPPANHSFPIDRYIMEFRV
AERWELLDDGIPGTEGEFFAKDLSQDTWYEFVRLAVMQDLIGEPSNIAGVSSTDIFPQPD
LTEDGLARPVLGIVATICFLAAAILFSTLAACFVNKQRKRLKRKKDPLSITHCRKSL
ESPLSSGKVSPESIRTLRAPSESSDDQGQPAKRMLSPTREKELSLYKTKRAISSKYS
VAKAEAEAEATTPIELISRGPDGRFVMDPVEMEPSLKSRRIEGFPPAEETDMYPEFRQSD
EENEDPLVPTSVAALKSQLTPLSSSQESYLPPPAYSPPRQPRGLEPGGLEGRQLQATGQA
RPPAPRPFHHGQYYGYLSSSSPGEVEPPFPYVPEVGSPLSSVMSSPPLPTEGPFHGHTIP
EENGENASNSTLPLTQTPTGGRSPPEWGRPEFPFGGLETPAMMFPHQLPPCDVPESLQPK
AGLPRGLPPTSLQVPAAYPGILSLEAPKGWAGKSPGRGPVPAPPAKWQDRPMQPLVSQ
QLRHTSQGMGIPVLPYPEPAEPGAHGGPSTFGLDTRWYEPQPRPRPSRQARRAEPSSLHQ
VVLQPSRLSPLTQSPLSSRTGSPELAARARPRPGLLQQAEMSEITLQPPAAVSFSRKSTP
STGSPSQSSRSRGSYRPAAGFTTLATGYPSPPPGPAPAGPGDSLDFVFGQTSPRRTGEE
LLRPETPPPTLPTSGKLQRDRPAPATSPPERALS
```

A search of sequence databases reveals that the NOV45b amino acid sequence has 425 of 911 amino acid residues (46%) identical to, and 553 of 911 amino acid residues (60%) similar to, the 1189 amino acid residue ptnr:SPTREMBL-ACC:Q9P2J2 protein from Homo sapiens (Human) (KIAA1355 PROTEIN). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV45b is expressed in at least prostate, brain (cerebellum). This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV45a polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 45E.

Table 45E. BLAST results for NOV45					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 5689397 dbj BAA82982.1 (AB028953)	KIAA1030 protein [Homo sapiens]	763	761/763 (99%)	761/763 (99%)	0.0
gi 15311046 ref XP_027486.2 (XM_027486)	KIAA1030 protein [Homo sapiens]	747	583/627 (92%)	583/627 (92%)	0.0
gi 18578690 ref XP_062186.2 (XM_062186)	similar to KIAA1355 protein [Homo sapiens]	904	413/472 (87%)	417/472 (87%)	0.0
gi 7243091 dbj BAA92593.1 (AB037776)	KIAA1355 protein [Homo sapiens]	1189	410/882 (46%)	527/882 (59%)	e-179
gi 18426807 ref NP_291086.1 (NM_033608)	neural cell adhesion molecule (Ncam)-like; KIAA1355 hypothetical protein (human); NCAM-like protein NRT1 [Mus musculus]	1179	409/866 (47%)	525/866 (60%)	e-177

Tables 45F-H list the domain descriptions from DOMAIN analysis results against NOV . This indicates that the NOV sequence has properties similar to those of other proteins known to contain this domain.

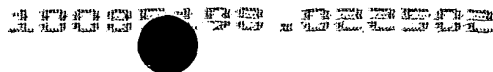
5

Table 45F. Domain Analysis of NOV45
gnl Smart smart00409 , IG, Immunoglobulin
CD-Length = 86 residues, 100.0% aligned
Score = 62.4 bits (150), Expect = 2e-10

Table 45G. Domain Analysis of NOV45
gnl Smart smart00408 , IGc2, Immunoglobulin C-2 Type
CD-Length = 63 residues, 92.1% aligned
Score = 54.7 bits (130), Expect = 4e-08

Table 45H. Domain Analysis of NOV45
gnl Pfam pfam00041 , fn3, Fibronectin type III domain.
CD-Length = 86 residues, 90.7% aligned
Score = 43.5 bits (101), Expect = 8e-05

The basic structure of immunoglobulin (Ig) molecules is a tetramer of two light chains and two heavy chains linked by disulfide bonds. There are two types of light chains: kappa and



lambda, each composed of a constant domain (CL) and a variable domain (VL). There are five types of heavy chains: alpha, delta, epsilon, gamma and mu, all consisting of a variable domain (VH) and three (in alpha, delta and gamma) or four (in epsilon and mu) constant domains (CH1 to CH4). The major histocompatibility complex (MHC) molecules are made of two chains. In class I the alpha chain is composed of three extracellular domains, a transmembrane region and a cytoplasmic tail. The beta chain (beta-2- microglobulin) is composed of a single extracellular domain. In class II, both the alpha and the beta chains are composed of two extracellular domains, a transmembrane region and a cytoplasmic tail. It is known that the Ig constant chain domains and a single extracellular domain in each type of MHC chains are related. These homologous domains are approximately one hundred amino acids long and include a conserved intradomain disulfide bond. Members of the immunoglobulin superfamily are found in hundreds of proteins of different functions. Examples include antibodies, the giant muscle kinase titin and receptor tyrosine kinases. Immunoglobulin-like domains may be involved in protein-protein and protein-ligand interactions.

Fibronectins are multi-domain glycoproteins found in a soluble form in plasma, and in an insoluble form in loose connective tissue and basement membranes. They contain multiple copies of 3 repeat regions (types I, II and III), which bind to a variety of substances including heparin, collagen, DNA, actin, fibrin and fibronectin receptors on cell surfaces. The wide variety of these substances means that fibronectins are involved in a number of important functions: e.g., wound healing; cell adhesion; blood coagulation; cell differentiation and migration; maintenance of the cellular cytoskeleton; and tumour metastasis . The role of fibronectin in cell differentiation is demonstrated by the marked reduction in the expression of its gene when neoplastic transformation occurs. Cell attachment has been found to be mediated by the binding of the tetrapeptide RGDS to integrins on the cell surface , although related sequences can also display cell adhesion activity. Plasma fibronectin occurs as a dimer of 2 different subunits, linked together by 2 disulphide bonds near the C-terminus. The difference in the 2 chains occurs in the type III repeat region and is caused by alternative splicing of the mRNA from one gene . The observation that, in a given protein, an individual repeat of one of the 3 types (e.g., the first FnIII repeat) shows much less similarity to its subsequent tandem repeats within that protein than to its equivalent repeat between fibronectins from other species, has suggested that the repeating structure of fibronectin arose at an early stage of evolution. It also seems to suggest that the structure is subject to high selective pressure. The fibronectin type III repeat region is an approximately 100 amino acid

domain, different tandem repeats of which contain binding sites for DNA, heparin and the cell surface . The superfamily of sequences believed to contain FnIII repeats represents 45 different families, the majority of which are involved in cell surface binding in some manner, or are receptor protein tyrosine kinases, or cytokine receptors.

5 The disclosed NOV45 nucleic acid of the invention encoding a Ig/fibronectin domain-like protein includes the nucleic acid whose sequence is provided in Table 45A or 45C or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 45A or 45C while still encoding a protein that maintains its Ig/fibronectin domain-like activities and physiological
10 functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example,
15 modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 1 percent of the bases may be so changed.

20 The disclosed NOV45 protein of the invention includes the Ig/fibronectin domain-like protein whose sequence is provided in Table 45B or 45D. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 45B or 45D while still encoding a protein that maintains its Ig/fibronectin domain-like activities and physiological functions, or a functional fragment
25 thereof. In the mutant or variant protein, up to about 54 percent of the residues may be so changed.

 The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

 The above defined information for this invention suggests that this Ig/fibronectin
30 domain-like protein (NOV45) may function as a member of a "Ig/fibronectin domain family". Therefore, the NOV45 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic,

diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV45 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the Ig/fibronectin domain-like protein (NOV45) may be useful in gene therapy, and the Ig/fibronectin domain-like protein (NOV45) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, fertility, or other pathologies or conditions. The NOV45 nucleic acid encoding the Ig/fibronectin domain-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV45 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV45 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV45 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV46

A disclosed NOV46 nucleic acid of 1247 nucleotides (also referred to as CG57682-01) encoding a G2/mitotic-specific cyclin B2-like protein is shown in Table 46A. The start and stop codons are in bold letters.

Table 46A. NOV46 nucleotide sequence (SEQ ID NO:109).

ATCCACATTGCATTTAGTTGTCAGGTATGCTATTAGGACTCTCCAAGAGACAGAACCAATAGGGGAAAT
GGGTATATGCAGAGACAGAGAGACAGGAGTTAATTCTAAACCTAAGAGTCATGTGACTATTAGGCATTCC

ATTTTAGAAAAAATTGGAATAGAGTTACAGCCAGAGCAGCACAAAGTAGCTAAGAAAGCTCAGAACACAC
AAAGTGCCAGTTCAACCCAGGGAACAACAAATGTCAACAAACAAGTAAACCTACTGCTTCTGTGAAAC
AGTACAGATGGAATGTTGGCTCCAAAGGGTCCTTCTCCCATACCTGAGGATGCCTCCGTGAAAGAAGAG
AACATCTGCCAGGCTTTTTCTGATGCTTTGCTCTACAAAAATTGAGGATATTGATAACAAAGATTGGAATA
ACCTCAGCTCTGCAGTAGCTATTAAAGGAAGGTATCTATCAGTACCTCAGGCAGCTGGAGATTTTGCA
GTTCATAAACCCACATGTCTTAGGTGAGGAGATGTAATGGACATAAAGCATACCATCCTGCTAGACTGG
TTGGTGCAAATCCACTCCAAGTTTAGGCTTCTTCAGGAGACTCTGTATGTGTGTGTGCCATTATGGATG
GATCTTTACTGTTTCAGCCAGTTTCCAGAGGAAGCTTCAACTAGTTTGGATTACTGCTCTGCTCTTGGC
TTCCAAGTATGAGGAGATGTTTTCTCCAAATACTGAAGGCTTTGTTTACATCAGACAAATGCTTTATACT
AGTTTCCAAATCCAGAAATGGAACTCTAATTTTGAAGAAAGCTGAAATTTGAGGTGGGTGGACCTTGC
CAGTACATCTTCTAAGGCAGCATCAAAAGCCGGAAGGCTGATGTTGAACAGCACACTTTAGCCAAATA
TTTGATGGAGCTGACTCTCATTGACTACGATATGATGCATTATCATCCTTCTAAGGCAGCAACAGCTGCT
TCCTGCTTGTCTCAGAAGGTTCTGGGCCAAGGAAATGGAACTTAAAGCAGCAGTGTTATACAGGATACA
CACAGAATGAAGTATTGGAAGTCATGCAGCACGTTGGCCAAAAATGTGCTGAAAGTAAATGAAAACCTTAAC
TAAATTATCGCCATCAAGAATAAGTATGCAAGCAACAAATTCCTGAAGATCAGCATGATCCCTCAGCTG
AACTCAAAGCCATCAAAGACCTTGCTTCCCTCTGATGGGAGGTCCTAGGCTGCA

ID:HSM800659|acc:AL080146.1 mRNA from Homo sapiens (Homo sapiens mRNA; cDNA

5 DKFZp434B174 (from clone DKFZp434B174); complete cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV46 polypeptide (SEQ ID NO:110) encoded by SEQ ID NO:109 has 404 amino acid residues and is presented in Table 46B using the one-letter amino acid code.

Signal P, Psort and/or Hydropathy results predict that NOV46 has a signal peptide and is
10 likely to be localized in the cytoplasm with a certainty of 0.6500.

Table 46B. Encoded NOV46 protein sequence (SEQ ID NO:110).

MLLGLSKETEPIGEMGICRDRETGVNSKPKSHVTIRHSILEKIGNRVATARAAQVAKKAQN
TQSASSTQGNNKCQQTETETFCFETVQMEMLAPKGPSPIPEDASVKEENICRAFSDALLY
KIEDIDNKDWNPNQLCSDYLKRGYQYLRQLEILQFINPHVLGGGDVNGHKHTILVDWL
QTHSKFRLLQETLYVCVAIMDGSLLVQPVSRKLQLVWITALLASKYEEMFSPNTEGFV
YITDNAYTSFQIQEMETLILKELKFVEGGPLPLHFLRQASKAGKADVEQHTLAKYLMELT
LIDYDMMHYHPSKAATAASCLSQKVLGGQKWNLQKQCYTGYTQNEVLEVMQHVAKNVLKV
NENLTKFIAIKNKYASNKFLKISMIPOLNSKAIKDLAFPLMGGS

A search of sequence databases reveals that the NOV46 amino acid sequence has 303 of 383 amino acid residues (79%) identical to, and 333 of 383 amino acid residues (86%) similar to, the 398 amino acid residue ptmr:SWISSNEW-ACC:O95067 protein from Homo sapiens (Human) (G2/MITOTIC-SPECIFIC CYCLIN B2). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV46 is expressed in at least Adrenal gland, Aorta, B-cells, Blood, Bone, Brain, Breast, CNS, Colon, Ear, Esophagus, Eye, Gall bladder, Germ Cell, Head and neck, Heart, Kidney, Larynx, Liver, Lung, Lymph, Marrow, Muscle, Neural, Omentum, Ovary, Pancreas,

Parathyroid, Peripheral nervous system, Placenta, Pooled, Prostate, Skin, Small intestine, Spleen, Stomach, Synovial membrane, Testis, Tissue culture, Tonsil, Uterus, Whole embryo, and adrenal gland. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV46 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 46C.

Table 46C. BLAST results for NOV46					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 4757930 ref NP_04692.1 (NM_004701)	cyclin B2 [Homo sapiens]	398	303/387 (78%)	333/387 (85%)	e-67
gi 5921730 sp O77689 CGB2 BOVIN	G2/mitotic-specific cyclin B2	398	285/385 (74%)	328/385 (85%)	e-62
gi 584914 sp P37883 CGB2 MESAUI	G2/mitotic-specific cyclin B2	397	288/385 (74%)	321/385 (82%)	e-58
gi 14198371 gb AAH08247.1 AAH08247 (BC008247)	Similar to cyclin B2 [Mus musculus]	398	278/382 (72%)	316/382 (81%)	e-51
gi 6680866 ref NP_031656.1 (NM_007630)	cyclin B2 [Mus musculus]	398	274/382 (71%)	313/382 (81%)	e-49

Tablez 46D-E list the domain descriptions from DOMAIN analysis results against NOV46. This indicates that the NOV46 sequence has properties similar to those of other proteins known to contain this domain.

Table 46D. Domain Analysis of NOV46
gnl Pfam pfam00134, cyclin, Cyclin, N-terminal domain. Cyclins regulate cyclin dependent kinases (CDKs). Cyclins contain two domains of similar all-alpha fold, of which this family corresponds with the N-terminal domain. CD-Length = 128 residues, 99.2% aligned Score = 119 bits (298), Expect = 3e-28

Table 46E. Domain Analysis of NOV46

gnl|Smart|smart00385, CYCLIN, domain present in cyclins, TFIIB and Retinoblastoma; A helical domain present in cyclins and TFIIB (twice) and Retinoblastoma (once). A protein recognition domain functioning in cell-cycle and transcription control.

CD-Length = 83 residues, 100.0% aligned

Score = 62.8 bits (151), Expect = 4e-11

Two B-type cyclins, B1 and B2, have been identified in mammals. Proliferating cells express both cyclins, which bind to and activate p34(cdc2). To test whether the two B-type cyclins have distinct roles, lines of transgenic mice were generated, one lacking cyclin B1 and the other lacking cyclin B2. Cyclin B1 proved to be an essential gene; no homozygous B1-null pups were born. In contrast, nullizygous B2 mice developed normally and did not display any obvious abnormalities. Both male and female cyclin B2-null mice were fertile, which was unexpected in view of the high levels and distinct patterns of expression of cyclin B2 during spermatogenesis. The expression of cyclin B1 overlaps the expression of cyclin B2 in the mature testis, but not vice versa. Cyclin B1 can be found both on intracellular membranes and free in the cytoplasm, in contrast to cyclin B2, which is membrane-associated. These observations suggest that cyclin B1 may compensate for the loss of cyclin B2 in the mutant mice, and implies that cyclin B1 is capable of targeting the p34(cdc2) kinase to the essential substrates of cyclin B2.

The disclosed NOV46 nucleic acid of the invention encoding a G2/mitotic-specific cyclin B2-like protein includes the nucleic acid whose sequence is provided in Table 46A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 46A while still encoding a protein that maintains its G2/mitotic-specific cyclin B2-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 9 percent of the bases may be so changed.

The disclosed NOV46 protein of the invention includes the G2/mitotic-specific cyclin B2-like protein whose sequence is provided in Table 46B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 46B while still encoding a protein that maintains its G2/mitotic-specific cyclin B2-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 21 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this G2/mitotic-specific cyclin B2-like protein (NOV46) may function as a member of a “G2/mitotic-specific cyclin B2 family”. Therefore, the NOV46 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV46 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the G2/mitotic-specific cyclin B2-like protein (NOV46) may be useful in gene therapy, and the G2/mitotic-specific cyclin B2-like protein (NOV46) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from Cardiomyopathy, Atherosclerosis, Hypertension, Congenital heart defects, Aortic stenosis, Atrial septal defect (ASD), Atrioventricular (A-V) canal defect, Ductus arteriosus, Pulmonary stenosis, Subaortic stenosis, Ventricular septal defect (VSD), valve diseases, Tuberous sclerosis, Scleroderma, Obesity, Transplantation, Adrenoleukodystrophy, Congenital Adrenal Hyperplasia, Hemophilia, Hypercoagulation, Idiopathic thrombocytopenic purpura, Immunodeficiencies, Graft versus host, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, or other pathologies or conditions. The NOV46 nucleic acid encoding the G2/mitotic-specific cyclin B2-like protein of the invention,

or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV46 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV46 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV46 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV47

A disclosed NOV47 nucleic acid of 15645 nucleotides (also referred to as CG57764-01) encoding a ALR-like protein is shown in Table 47A. The start and stop codons are in bold letters.

Table 47A. NOV47 nucleotide sequence (SEQ ID NO:111).

ATGTCCCCTCCACCTGAAGAGTCACCCATGTCTCCACCACCGAGGCATCTCGTCTGTTCCACCATTTGAAGAGTCTCCTCTGTCCCCTCCACCTGAGGAGTCTCCCCTTTCCCCACCACTGAGGCATACGCCTGTCCCCACCACTGAGGACTCGCCTATGTCCCCACCACTGAA GAATCACCTATGTCCCCCCCACCTGAGGTATCGCGCCTATCCCCCTGCCTGTGGTGTCA CGCTGTCTCCACCGCCTGAGGAATCTCCCTTGTCCTCCACCGCCTGAGGAGTCTCCACG TCCCCTCCACCTGAGGCTTCACGCCTCTCCCCACCACTGAGGACTCCCCACATCCCCA CCACCTGAGGACTCACCTGCTTCCCCACCAACCGAGGACTCGCTCATGTCCCTGCCGCTG GAGGAGTCACCCCTGTTGCCACTACCTGAGGAGCCGCAACTCTGCCCCCGGTCCGAGGGG CCGCACCTGTACCCCGCCTGAGGAGCCGCACCTGTCCCCCGGCCTGAGGAGCCACAC CTATCTCCGAGGCTGAGGAGCCACACCTGTCCCCCAGCCTGAGGAGCCATGCCTATGC GCTGTGCCTGAGGAGCCACACTTGTCCCCCAGGCTGAGGGACCACATCTGTCCCCTCAG CCTGAGGAATTGCACCTGTCCCCCAGACTGAGGAGCCGCACCTGTCTCTGTGCCTGAG GAGCCATGCTTGTCCCCCAACCTGAGGAATCACACCTGTCCCCCAGTCTGAGGAGCCA TGCTGTCCCCCGGCCTGAGGAATCGCATCTGTCCCCTGAGCTTGAGAAGCCACCCCTG TCCCCTCGGCCTGAAAAGCCCCCTGAGGAGCCAGGCCAATGCCCTGCACCTGAGGAGCTG CCCTGTTCCTCCCCCTGGGGAACCATCCTTATCTCCCTTGCTTGAGAGCCAGCCCTG TCTGAGCCTGGGGAACCACTCTGTCCCCTTGCCCCGAGGAGCTGCCGTTGTCCCCATCT GGGGAGCCATCCTTGTGCGCTCAGCTGATGCCACCAGATCCCCTTCCTCCTCCACTCTCA CCCATTATCACAGCTGCGGCCCCACCGGCCCTGTCTCCTTGGGGGAGTTAGAGTACCCC TTGGTGCCAAAGGGGACAGTGACCTGAGTCACCGTTGGCTGCCCCCATCCTGGAGACA CCCATCAGCCCTCCACCAGAAGCTAACTGCACTGACCTGAGCCTGTCCCCCTATGATC CTTCCCCCATCTCCAGGCTCCCCAGTGCGGGCCGGCTTCTCCCATCTGATGGAGCCCCTT CCTCCTCAGTGTTCGCCACTCCTCAGCATTCCTGGTTCCCCAAACTCCCCTCCTTCC CAGTGCTCTCCTCCTGCCCTACCACTGTCCGTTCCTCCTCCCGTTGAGTCCCATAGGGAAG GTAGTGGGGGTCTCAGATGAGGCTGAGCTGCACGAGATGGAGACTGAGAAAGTTTCAGAA CCTGAATGCCAGCCTTGAACCCAGTGCCACCACTCCTCTCCCTTCCCCAATGGGGGAC CTTTCTGCCCCGCCCCAGCCCTGCCCCAGCCCTGGATGACTTCTCTGGCCTAGGGGAA GACACAGCCCTCTGGATGGGATTGATGCTCCGGGTTACAGCCAGAGCCTGGACAGACC CTGGCAGTTTGGCTAGTGAACCTAAAGGCTCCCTGTGCTCCTGGACCCCGAGGAGCTG GCCCTGTGACCCCTATGGAGGTCTACCCGAATGCAAGCAGACAGCAGGGCGGGGCTCA CCATGTGAAGAACAGGAAGAGCCACGTGCACCGGTGGCCCCCACACCCCACTCTCATC AAATCCGACATCGTTAACGAGATCTTAATCTGAGCCAGGGTGATGCCAGTGCCAGTTT CCTGGCTCAGAGCCCTCCTGGGCTCTCCAGACCCGAGGGGGTGCTCCCTGTCCATG

GAGTTGGGGTCTCTACGGATGTTAGTCCAGCCGAGATGAGGGCTCCCTACGGCTCTGT
 ACTGACTCACTGCCAGAGACTGATGACTCACTATGTGCGATGCTGGGACAGCTATCAGC
 GGAGGCAAAGCTGAGGGGAGAAGGGGCGGCGCAGCTCCCCAGCCGTTCCCGCATC
 AAACAGGGTCGAGCAGCAGTTCCTCCAGGAAGACGCCGGCTCGTGGAGGAGCCCATGGA
 GGGCGTGGTAGAGGACGGGCGGGCTAAAGTCAACTGCTTCTTCCATTGAGACTCTGGTA
 GTTGCTGACATTGATAGCTCTCCAGTAAGGAGGAGGAGGAAGAAGATGATGACACCATG
 CAGAATACCGTGGTTCTCTTCTCCAAACACAGACAAATTTGTCTAATGCAGGACATGTGT
 GTGGTATGTGGCAGCTTTGGCCGGGGGCGAGAGGGCCACCTCCTTGCCCTGTTCCGAGTGC
 TCTCAGTGCTATACCCCTTACTGTGTCAACAGCAAGATCACCAGGTGATGCTGTCAAG
 GGCTGGCGTTGTGTGGAGTGTATTGTGTGTGAGGTGTGTGGCCAGGCCTCCGACCCTCA
 CGCCTGTCTGTGTGATGACTGTGATATTAGCTACCACACATACTGCCTGGACCCCCA
 CTGCTCACCGTCCCCAAGGGCGGCTGGAAGTGCAAGTGGTGTGTCTGTATGCAGTGT
 GGGGTGTCTTCCCTGGCTTCCACTGTGAATGGCAGAATAGTTACACACACTGTGGGCC
 TGTGCCAGCCTGGTGACCTGCCCTATCTGTATGCTCCTTACGTAGAAGAGGACCTACTA
 ATCCAGTCCCGCACTGTGAACGGTGGATGCATGCAGGCTGTGAGAGCCTCTTACAGAG
 GACGATGTGGACACGCACCCGATGAAGGCTTTGACTGTGTCTCTGCCAGCCCTACGTG
 GTAAAGCCTGTGGCGCTGTGACCTCCAGAGCTGGTGCCCATGAAGGTGAAAGAGCCA
 GAGCCCCAGTACTTTTCGCTTCGAAGGCGTGTGGCTGACAGAACTGGCATGGCCTTGCTG
 CGTAACCTGACCATGTCAACACTGCACAAGCGGCGCAACGGCGAGGACGGCTTGGCCTC
 CCAGGCGAGGCGAGGATTGGAGGGTTCTGAGCCCTCAGATGCCCTTGGCCCTGATGACAAG
 AAGGATGGGGACCTGGACACCGATGAGCTGCTCAAGGTGAAGGTGGTGTGGAGCACATG
 GAGTGCAGAAATAAACTGGAGGGCCCCGTGAGCCCTGATGTGGAGCCTGGCAAAGAGGAG
 ACCGAGGAAAGCAAAAACGCAAGCGTAAACCATATCGGCCTGGCATTGGTGGTTTCATG
 GTGCCAGACGCGAAATCCACACACGCACGAAAAGGGGCTGCTGCACAGGCGGAGGTG
 TTGAGTGGGGATGGGACGCCGACGAGGTGATACCTGCTGACCTGCCCTGCAGAGGGCGCC
 GTGGAGCAGAGCTTAGCTGAAGGGGATGAGAAGAAGAAGCAACAGCGCGAGGGCGCAAG
 AGGAGCAAACCTGGAGGGCATGTTCCCTGCTTACTTGCAGGAAGCCTTCTTGGGAAGGAG
 CTGCTGGACCTGAGCCGTAAGGCCCTTTTTCAGTTGGGGTGGGCGGCCAAGCTTTGGA
 CTAGGGACCCCAAAGACCAAGGGAGATGGAGGCTCAGAAAGGAAGGAACCTCCACATCG
 CAGAAAGGAGATGATGGTCCAGATATTGCAGATGAAGAATCCCGTGGCCTCGAGGGCAAA
 GCCGATACACCAGGACCTGAGGATGGGGGCGTGAAGGCATCCCCAGTGCCAGTGACCCCT
 GAGAAGCCAGGCACCCAGGTGAAGGGATGCTTAGCTCTGACTTAGACAGGATTTCCACA
 GAAGAACTGCCCAAGATGGAATCCAAGGACCTGCAGCAGCTCTTCAAGGATGTTCTGGGC
 TCTGAACGAGAACAGCATCTGGGTTGTGGAACCCCTGGCCTAGAAGGCAGCCGTACGCCA
 CTGCAGAGGCCCTTTCTTCAAGGTGGACTCCCTTTGGGCAATCTGCCCTCCAGCAGCCCA
 ATGGACTCTACCCAGGCCTCTGCCAGTCCCCGTTCCTGGATTCTAGGGAGCGCGGGGGC
 TTTCTTGAAGCCGGAACCCGTGAGCCCGACAGCCCTGGACGGGCTCAGGTGGCACCAAG
 CCCTCCACCCCAACAACCCCAACCGAGGGTGAGGGCGACGGACTCTCCTATAACCAAG
 CGGAGTCTTCAGCGCTGGGAGAAGGATGAGGAGTTGGGCCAGCTGTCCACCATCTGCCT
 GTGCTCTATGCCAACATTAATTTTCTAATCTCAAGCAAGACTACCCAGACTGGTCAAGC
 CGTTGCAAAACAAATCATGAAGCTCTGGAGAAAGGTTCCAGCAGCTGACAAAGCCCCCTAC
 CTGCAAAAGGCCAAAGATAACCGGGCAGCTCACCGCATCAACAAGGTGCAGAAGCAGGCT
 GAGAGCCAGATCAACAAGCAGACCAAGGTGGGCGACATAGCCCGTAAGACTGACCGACCG
 GCCCTACATCTCCGATTTCCCCCGCAGCCAGGGGCACTGGGCAGCCCCCCCCCGCTGCT
 GCCCCACCATTTTCATTGGCAGCCCCACTACCCCGCCGGCTTGTCTACCTCTGCGGAC
 GGGTTCTGAAGCCGCGCGGGCTCGGTGCCCTGGCCCTGACTCGCCTGGTGAGCTCTTC
 CTCAAGCTCCCAACCCAGGTGCCCGCCCAAGCGCTTCGAGGACCCCTTTGGACTGGCC
 CCTGCTATCCCTGGAGCCCCGCTTCCCCACGGCACCGCCACCTATCCCCCTATCCT
 AGTCTACGGGGGCCCTGCGCAGCCCCGATGCTGGGCGCTCATCTCGTCTGGGGCT
 GGCCAGCCAGGGGAATTCACACTACCCACCTGGCACCCCAAGACACCAAGCCCTCCACA
 CCTGACCCGTTCTCAAACCCCGCTGCCCTCGCTGATAACTTGGCTGTGCCCTGAGAGC
 CCTGGGGTAGGGGAGGCAAAGCTTCCGAGCCCTGCTCTCGCCCCACCTTTTGGGGAG
 TCCCGGAAGGCCCTAGAGGTGAAGAAGGAAGAGCTTGGGGCATCTCTCTAGCTATGGG
 CCCCCAAACCTGGGCTTTGTGACTCACCTCCTCAGGCACCCACCTGGGTGGCCTGGAG
 TTAAGACACCTGATGTCTTCAAAGCCCCCTGACCCCTCGGGCATCTCAGGTAGAGCCC
 CAGAGCCCGGGCTTGGGCTAAGGCCCCAGGAGCCACCCCTGCCAGGCTTTGGCACCT
 TCTCTCCAAGTACCCAGACATCTTTCGCCCTGGCTCCTACACTGACCCATATGCTCAG
 CCCCCATTGACTCCTCGGCCCAACCTCCGCCCTGAGAGCTGCTGTGCTCTGCCCCCT
 CGCTACTGCCCTCCGACCTTTCTCCGAGTGCTGTGCTCCTCAGTCCAGTCCAGTCCAGC
 TCCAGTCTCCACTGACACCCCGGCTCTGTCTGCTGAAGCTTTTGGCCATCACCCGTT
 ACCCTCGCTTCCAGTCCCTGACCTTATTCTCGCCACCCCTCACGCCCTCAGTCCCGT
 GACCCATTTGCCCCATTGCATAAGCCACCCGACCCAGCCCCCTGAAGTTGCCTTTAAG
 GCTGGGTCTTAGCCCACTTCTGTTGGGGCTGGGGGGTCCCAGCAGCCCTGCCCGCG
 GGGCCAGCGGTGAGCTCCATGCCAAGGTCCCAAGTGGGCGAGCCCCCAATTTTGTCCGG
 TCCCTGGGACGGGTGCATTTGTGGGCACCCCTCTCCCATGCGTTTCACTTTCCCTCAG
 GCAGTAGGGGAGCCTTCCCTAAAGCCCCCTGTCCCTCAGCTGGTCTCCCGCCACCCCAT
 GGGATCAACAGCCATTTTGGGCCCGGCCCCACCTTGGGCAAGCCTCAAAGCACAACTAC
 ACAGTAGCCACAGGGAACCTTCAACCATCGGGCAGCCCCCTGGGGCCAGCAGCGGGTCC
 ACAGGGGAGAGCTATGGGCTGTCCCACTACGCCCTCCGTGGTTCTGCCACCACTGCA

CCCAGCGGATCCCTCCCCTACCTGTCCCATGGAGCCTCACAGCGATCAGGCATCACCTCT
CCTGTGAAAAAGCGAGAAGACCCAGGGACTGGATGGGTAGCTCTTTGGCGACAGTGAA
CTCCCAGGTACCCAGGACCCAGGCATGTCCGGCCTTAGCCAAACAGAGCTGGAGAAGCAA
CGGCAGCGCCAGCGGCTACGAGAGCTGCTGATTGGGCAGCAGATCCAGCGCAACACCCCTG
CGGCAGGAGAAGGAAACAGCTGCAGCAGCTGCGGGAGCAGTGGGGCCTCCAGGCAGCTGG
GGTGTGAGCCACAGCAGCCCTGCCTTTGAGCAGCTGAGTCGAGGCCAGACCCCTTTGCT
GGGACACAGGACAAGAGCAGCCTTTGTTGGGTTGCCCCAAGCAAGCTGAGTGGCCCCATC
CTGGGGCCAGGGTCTTCCCTAGCGATGACCGACTCTCCCGGCCACCTCCACCAGCCACG
CCTTCTCTATGATGTGAACAGCCGCAACTGGTAGGAGGCTCCCAAGCTTTCTATCAG
CGAGCACCCATCTCTGGGTCCCTGCCCTTACAGCAGCAACAGCAACAACTGTGGCAGCAA
CAACAGGCAACAGCAGCAACCTCCATGCGATTTGCCATGTGAGCTCGCTTTCCATCAACT
CCTGGACCTGAACCTGGCCGCCAAGCCCTAGGTTCCCCGTGGCGGAATTTCCACCCGT
CTGCCAGGCCCTGGTGAGCCAGTGCCCTGGTCCAGCTGGTCTGCCAGTTCATTGAGCTG
CGGCACAATGTACAGAAAGGACTGGGACCTGGGGGCACTCCGTTTCTGGTCAAGGCCCA
CCTCAGAGACCCCGTTTTTACCCTGTAAAGTGAGGACCCCCACCGACTGGCTCCTGAAGG
CTTCGGGGCCTGGCGGTATCAGGTCTTCCCCACAGAAACCTCAGCCCCACCGGCCCT
GAATTGAACAACAGTCTTCAATCAACACCCACACCAAGGGTCTACCTGCGCAACTGGT
TTGGAGCTGGTCAACCGGCCCTCGAGCAGTGGCTGGCGGCCCAATCCTCTGGCC
CTGGAAGCTGGGAAGTTGCCCTGTGAGGATCCCGAGCTGGATGACGATTTTGATGCCAC
AAGGCCCTAGAGGATGATGAAGAGCTTGCTCACCTGGGTCTGGGTGTGGATGTGGCCAA
GGTGTGATGAACCTGGCACCTTAGAAAACCTGGAGACCAATGACCCCCACTGGATGAC
CTGCTCAATGGAGACGAGTTGACCTGCTGGCATATACTGATCCTGAGCTGGACACTGGG
GACAAGAAGGATATCTTCAATGAGCACCTGAGGCTGGTAGAATCGGCTAATGAGGAGGCT
GAACGGGAGGCCCTGTGCGGGGGTGGAGCCAGGACCTTGGGGCCTGAGGAGCGCCCT
CCCCCTGTGCTGATGCCTCTGAACCCCGCCTGGCATCTGTGCTCCCTGAGGTGAAGCCC
AAGGTGGAGGAGGGTGGAGCCACCCCTCTCCTTGCCAATTCAACATTGCTACCCCCAAG
GTAGAGCCCGCACCTGTGCCAATTCCTTGGCTGGGGCTAAAGCCAGGACAGAGCATG
ATGGGCAGCCGGGATACCCGATGGGCACAGGGCCATTTTCTAGCAGTGGGCACACAGCT
GAGAAGGCCCTCCTTTGGGGCCACGGGAGGGCCACCAGCTCACCTGCTGACCCCCAGCCA
CTGAGTGGCCAGGAGGATCCTCCCTGCTGGAAGTTTGGAGCTCGAGAGTGGGGCTTTG
ACCTTGCCCTGGTGGACCTGAGCATCTGGGGATGAGCTAGACAAGATGGAGAGCTCACTG
GTAGCCAGCGAGTTACCCCTGCTCATTGAGGACCTGTTGGAGCATGAGAAGAAGGAGCTG
CAACGCTGGCTCCATCCATGGCTATGGTGTCCAATCAAGGGCATATGCTAAGTGGGCAG
CATGGAGGGCAGGCAGGCTTGGTACCCAGCAGAGCTCACAGCCAGTGTATCACAGAAG
CCCATGGGCACCATGCCACCTTCCATGTGCATGAAGCCGACGAATTGGCAATGCAGCAG
CAGCTGGCAAAACAGCTTCTTCCAGATACAGACCTGGACAAATTTGCTGCAGAAGATATC
ATTGGTCCCATTGCAAAGGCCAAGATGGTGGCTTTGAAAGGCATCAAGAAAAGTGATGGCT
CAGGCGAGCATTGGGGTGGCACCTGGTATGAACAGACAGCAAGTGTCTCTGCTAGCCAG
AGGCTCTCGGGGGACCTAGCAGTGATCTGCAGAACCATGTGGCAGCTGGGAGTGGCCAG
GAGCGGAGTGCTGGTGTATCCCTCCCAGCCTCGTCCCAACCCGCCCACTTTTGCTCAGGGA
GTGATCAATGAAGCTGACCAGCGGCAGTATGAGGAGTGGCTGTTCCATACCCAGCAGCTC
CTACAGATGCAGCTGAAGGTGCTAGAGGAGCAGATTGGTGTACACCGCAAGTCCCGGAAG
GCTCTGTGTGCCAAGCAGCGCACTGCCAAAAAGCTGGCCGTGAGTTCCAGAAAGCTGAT
GCTGAGAAGCTCAAGCTGGTTACAGAGCAGCAGAGCAAGATCCAGAAACAACCTGGATCAG
GTCCGGAAACAGCAGAAGGAGCACACTAATCTCATGGCAGAATATCGGAACAAGCAGCAG
CAACAACAGCAGCAGCAGCAACAACAGCAACAGCACTCAGCTGTGCTGGCTCTCAGC
CCTTCCAGAGTCCCCGGCTGCTCACCAAGCTCCCTGGTCAAGCTGCTCCCTGGCCATGGG
CTGCAGCCACCACAGGGGCTCCGGGTGGGCAAGCCGAGGTCTTCGCTGACCCCTGGG
GGTATGGCACTACCTGGACAGCCTGGTGGCCCCCTTCTTAATACAGCTCTGGCCCAACAG
CAGCAACAGCAACATTCTGGTGGGGCTGGATCCCTGGCTGGCCCTTTCAGGGGGCTCTTC
CCTGGCAACCTTGCTCTTGAAGCCTCGGACCTGATTCAAGGCTTTTACAGGAAAGGCAG
CTGCAGCTGCAGCAGCAACGTATGCAGCTGGCCAGAACTGCAGCAGCAGCAGCAGCAG
CAACAGCAGCAGCAGCACCTTCTAGGACAGGTGGCAATCCAGCAGCAACAGCAGCAGGGT
CCTGGAGTACAGACAAACCAAGCTCTGGGTCCCAAGCCCCAGGGCCTTATGCCTCCCAGC
AGCCACCAAGGCCTCCTGGTCCAGCAGCTGTCCCTCAACCAACCCAGGGGCCCGAGGC
ATGCTGGGGCCTGCCAGGTGGCTGTGTTGCAGCAGCAGCACCCCTGGAGCTTTGGGGCCC
CAGGGCCCTCACAGACAGGTGCTTATGACCCAGTCCCGGGTGTCTAGTTCCCCCAGCTG
GCACAGCAGGGTCAAGGGCCTTATGGGACACAGGCTGGTACAGCCAGCAGCAGCAGCAG
CAACAACAGCACCAACAGCAAGGTCATGGCAGGGCTGTCCCTCTTACAGCAAGTCTG
ATGTCACACAGTGGGCAGCCAACTGAGCGCTCAGCCCATGGGCTCTTACAGCAGCTT
CAGCAGCAGCAGCAGCTGCAACAGCAACAGCAACTTACAGCAGCAGCAGCAGCAGCTA
CAACAGCAACAGCAACTTACAGCAGCAACAGCTTCAACAGCAGCAACAGCAGCAGCAGCTT
CAACAACAGCAGCAGCAACAGCTTCAACAGCAGCAACAGCAGCTACAACAGCAACAGCAA
CAACAACAGCAGCAGCTTCAACAGCAGCAGCAACAGCAGCAGATGGGCTTTTAAACAG
AGTCGAACCTTTACTGTCCCTCAGCAACAACAGCAGCAGCAAGTGGCACTTGGCCCTGGC

ATGCCAGCAAAGCCTCTTCAACACTTTTCTAGCCCTGGAGCCCTGGGTCCAACCCCTCCTC
CTGACGGGCAAGGAACAAAACACCGTAGACCCAGCCGTTTCTTACAGAGGCACTGAGGGG
CCCTCTACACATCAGGGAGGGCCGTTAGCAATAGGAACCTACCCCTGAGTCAATGGCCACT
GAACCAGGAGAGGTAAAGCCCTCACTCTCTGGGGACTCACAACTCCTGCTTGTCCAACCC
CAGCCCCAGCCTCAGCCAGCTCTCTGCAGCTGCAGCCACCTCTGAGGCTTCCAGGACAA
CAGCAGCAGCAAGTTAGCCTGCTCCACACAGCAGGTGGAGGAAGCCATGGGCAGCTAGGC
AGTGGATCATCTTCTGAGGCCTCATCTGTGCCCCACCTGCTGGCTCAGCCCTCTGTTTCC
TTAGGGGATCAGCCTGGGTCCATGACCCAGAACCTTCTGGGCCCCAACAGCCCATGCTA
GAGCGGCCCATGCAAAATAATACAGGGCCACAACCTCCCAAACAGGACCTGTCTCCAG
TCTGGGAGGGTCTGCCTGGGGTTGGAATCATGCCTACGGTGGGTGAGCTTCGAGCACAG
CTCCAAGGAGTCTGGCCAAAAACCCACAGCTGCGGCACCTTAAGTCTCAGCAGCAGCAG
CAGCTACAGGCACCTCTCATGCAGCGGCAGCTGCAGCAGAGTCAGGCAGTACGCCAGACC
CCACCTTACCAGGAGCCTGGGACCCAGACCTCTCCCTCCAGGGCCTCCTGGGCTGCCAA
CCTCAACTTGGGGGCTTCCCTGGACCAGACAGGCCCCCTCCAGGAGCTAGGGGAGGG
CCTCGACCTCAGGGCCACCCCGGCTCCTGCCCCACCAGGAGCCTTATCTACAGGACCA
GTCCTTGGCCCTGTCCATCCACACCTCCACCATCCAGCCCTCAAGAGCCAAAGAGACCT
TCACAATTACCTTCCCCAGCTCCAGCTTCCACTGAGGCCAGCTCCCTCCACCCAT
CCAGGGACCCCCAAACCTCAGGGGCCAACCTTGGAGCCGCTCCTGGGAGGGTCTCACCT
GCTGCTGCCAGCTTGCAGATACCTTGTTTAGCAAGGGTCTGGGACCTTGGGATCCCCCA
GACAACCTAGCAGAAACCCAGAAGCCAGAGCAGAGCAGCCTGGTACCTGGGCATCTGGAC
CAGCTGAATGGACAGGTGGTGCCTGAGGCATCCAACTCAGCATCAAGCAGGAACCTCGG
GAAGAGCCATGTGCCCTGGGAGCCAGTCAGTGAAGAGGGAGGCCAATGGGGAGCCAATA
GGGGCACCAGGAACCAGCAACACCTCCTGCTGGCAGGCCCTCGCTCAGAAGCTGGGCAT
CTGCTCTTGCAGAAGCTACTCCGGGCAAGAATGTGCAACTCAGCACTGGGCAGGGGTCC
GAGGGGCTGCGAGCTGAGATCAACGGGCACATTGACAGCAAGCTGGCTGGGCTGGAGCAG
AAACTACAGGGTACCCCCAGCAACAAGGAGGATGCAGCAGCAAGGAAGCCTTTGACACCG
AAGCCCAAGCGGGTACAGAAGGCAAGCGACAGGTTGGTGAAGCTCCCGAAAGAAGCTGCGG
AAGGAGGACGGGCTCAGGGCCAGCGAGGCCTTGTGAAACAGCTGAAACAGGAGCTGTCC
CTGCTGCCCTTAACGGAGCCTGCTATCACCGCCAATTTTAGCCTCTTGGCCCCCTTGGC
AGTGGCTGCCAGTCAATGGGCAGAGCCAGCTGAGGGGGGCTTTGGAAGTGGGGCGCTG
CCCACTGGCCCTGACTACTATTCCAGCTGCTTACCAAGAATAACCTGAGTAACCCGCCG
ACACCACCTCGTCGCTGCCCCCCACCCACCCCATCGGTGCAGCAGAAGATGGTGAAT
GGCGTCACCCCATCTGAAGAGCTGGGGGAGCACCCCCAAGGATGCTGCCTCTGCCCCGGAT
AGTGAAGGGCACTGAGGGTACTTTCAGAGGTGAAGAGTCTAGACCTGCTGGCTGCCTTG
CCTACACCCCTCACAATCAGACTGAGGATGTGAGGATGGAGAGTATGAGGATAGCGAT
TCTCCTGACAGCATTGTGCCAGCTTCTATCCCTGAGAGCATCTTGGGGGAGGAGGCCCT
CGTTTTCCCTCATCTGGGCTCAGGCCGTGGGAGCAAGAGGACCGGGCCCTCTCCCTGTG
ATCCCCCTCATTCTCGGGACAGCATCCCACTTCCAGATACCAAACCTTATGGGGCC
CTTGGCCTGGAGTCCCTGGAAAGCTGCCTGTCACAACTTGGGAAAAGGGCAAGGAAGT
GAGGTGTGAGTCACTGCTCACAGTCTCTGCTGCTGCAGACAAGAACCTGAATGGCGTGATG
GTGGCAGTGGCGGAGCTGCTGAGCATGAAGATCCCCAATCCTATGAGGTGCTGTTCCCA
GAGAGCCCCCGCCGGGAGGCACTGAGCCAAAGAAGGGGAAGCTGAGGGTCTTGGTGGG
AAGAAAAGGGTCTGGAAGGCAAGAGCCAGACACTGGCCCTGATTGGCTGAAGCAGTTT
GATGCACTGTTGGTGGCTATACCTGAAGAGGCAACTAGACATCTTGAGCCTCCTGAAA
CAGGAGAGCCCCGCCAGAGCCACCCACTCAGCACAGGTATACCTACAATGTCTCCAAT
CTGGATGTGCGACAGCTCTCGGCCCACTCTGGAAGAACCCTCCCGCCCCCTTCCCC
TTGGCACCTTCTCCTGCCAGTCCCCCTACTGAGCCCTTGGTTGAACCTCCACCGAACC
TTGGCTGAGCCACCGTCCCTCACCTCTGCCACTGGCCCTCATCCCTGAATCAGCCCGA
CCCAAGCCCCGTGCCCGCCCCCTGAAGAAGGTGAAGATACCCGTCTCTCGCCTCAAG
AAATGGAAGGAGTGCCTGGAAGCGGCGCTTACGAGGTGCCATGTGAGAGCTTTTGGT
GTGAACAGTCTGGAAGTAAATTTAGGACCAGAAGCGAGAATGGCGTTTAAATCCATATC
CAAGAAAGCAGCAATTACACTACTGTGAAGATTGAAGATGGCAAAGTATATTTTACATCC
GATGCAGGAATTGCTGGGAAAGTGGAGAGAAATATTCTGAAGTATATGTTGCAGACGGC
CACTGGCACACTTTTCTAATTGGGAAAATGGAACAGCAACAGTATGTCTGTGACAGA
ATATATAACAGAGATATTATCCACCTACTCAGGACTTCGGTGGCCTTGATGTGCTTACT
ATATCACTTGGAGGAATTCACCCCAATCAAGCACATCGAGATGCCCAAACAGCAGGTTT
GATGGCTGCATTGCTTCTATGTGGTATGGTGGAGAAAGTCTTCTTTCAGCGGGAAGCAT
AGCTTGGCCTCCATCTCAAAAACAGATCCCTCAGTGAAGATTGGCTGCCGTGGCCGAAC
ATTTGTGCCAGCAACCCCTGCTGGGGTGATTTGCTGTGCATTAATCAGTGGTATGCCTAC
CAGGTGTGCTCCCTCCTGGGACTGTGCTCCACCCGTGCCAGAATGGTGGCAGCTGTGAG
AGGCTGCTGCACTCCGGCTCACCTGTAGCTGCCAGACTCGCACCGGGAAGGACCTGT
GAGATGGTGGTGGCTGTCTTGGCGTCTGTCTCAGGGGAAGGTGTGCAAAGCTGGA
AGTCTGCGGGGATGTCTGTGTTCTGAGTCAGGGCCCTGAAGAGATCTCTGTGCTTTG
TGGGCTGTGCTGCCATCGTGGGCAGCTGCGCAACCGTCTTGGCCCTCCTGGTCTTAGC
CTGATCCTGTGTAAACAGTGCAGGGGAAGAAGGCCAAAAATCCCAAAGAGGAGAAGAAA
CCGAAGGAGAAGAAGAAAAAGGGAAGTGAGAACGTTGCTTTTGATGACCTGACAATATC
CCTCCCTATGGGGATGACATGACTGTGAGGAAGCAGCCTGAAGGGAACCCAAAACAGAT
ATCATTGAAAGGAAAACCCCTACCTTATCTATGATGAAACTGATATTCCTCACAACCTCA
GAAACCATCCCCAGCGCCCCCTTGGCATCTCCAGAGCAGGAGATAGAGCACTATGACATT

GACAACGCCAGCAGCATCGCCCCTTCGGATGCAGACATCATCAACACTACAAGCAGTTC
CGCAGCCACACACCAAAATTTTCAATCCAGAGGCACAGTCCCCTAGGCTTTGCAAGGCAA
TCCCCCATGCCCTTAGGAGCAAGCAGTTTGACTTACCAGCCTTCATATGGTCAAGGTTTG
AGAACCAGCTCCCTAAGCCACTCAGCATGCCCACTCCCAACCCTCTGTCTCGACACAGT
CCAGCCCCTTTCTCCAAATCTTCTACGTTCTATAGAAACAGCCCAGCAAGGGAATTGCAT
CTTCCTATAAGGGATGGTAATACTTTGGAAATGCATGGTGACACCTGCCAACCTGGCATT
TTCAACTATGCCACAAGGCTGGGAAGGAGAAGCAAGAGTCTCAGGCCATGGCATCACAT
GGTTCTAGACCAGGGAGTCGCCTAAAGCAGCCGATTGGGCAGATTCCACTGGAATCTTCT
CCTCCAGTCGGACTTTCTATTGAAGAAGTGGAGAGGCTCAACACACCTCGCCCTAGAAAC
CCAAGTATCTGCAGTGCAGACCATGGGAGGTCTTCTTCCAGAGGAGGACTGCAGAAGGCCA
CTGTCTAGAACAAGGAATCCAGCGGATGGCATTCAGCTCCAGAATCCTCTTCTGATAGT
GACTCCCATGAATCTTTCACTTGCTCAGAAATGGAATATGACAGGGAGAAGCCAATGGTA
TATACTTCCAGAATGCCCAAATTATCTCAAGTCAATGAATCTGATGCAGATGATGAAGAT
AATTATGGAGCCAGACTGAAGCCTCGAAGGTACCACGGTCGCAGGGCCGAGGGAGGACCT
GTGGGCACCCAGGCAGCAGCACCAGGCACTGCTGACAACACTGCCCATGAAGCTAGGG
CAGCAAGCAGGGACTTTCAACTGGGACAACCTTTGAACTGGGGCCCTGGCTTTGGCCAT
TATGTAGATGTTTTAAAGATTTGGCATCTCTTCCAGAAAAAGCAGCAGCAAATGAAGAA
GGCAAAGCTGGGACAACATAACCAGTCCCCAAAGATGGGGAAGCAGAACAGTATGTGTGA
AGTTTATGTACTGGCACTATAAAATATAAAAACAAGAAATAATACTTCAAACCATTGTAA
AGTTGCTGACTAGGTTGGGGTCACATTTGAAAAACAGGGCCAGTATGGGACTAGTGGGTG
GGAGGGGAAAACTTTTAAATATTAAACCACATTGCTGGCTGAAA

In a search of public sequence databases, the NOV47 nucleic acid sequence, located on chromosome 12q12.14 has 14307 of 14312 bases (99%) identical to a gb:GENBANK-ID:AF010404|acc:AF010404.1 mRNA from Homo sapiens (Homo sapiens ALR mRNA, complete cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV47 polypeptide (SEQ ID NO:112) encoded by SEQ ID NO:111 has 5159 amino acid residues and is presented in Table 47B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV47 has no signal peptide and is likely to be localized in the cytoplasm with a certainty of 0.9800.

Table 47B. Encoded NOV47 protein sequence (SEQ ID NO:112).

MSPPPEESPMSPPEASRLFPFPFEESPLSPPEESPLSPPEASRLSPPPEDSPMSPPPE
ESPMSPPEVSRLSPLPVVSRLSPPEESPLSPPEESPTSPPEASRLSPPPEDSPTSP
PPEDSPASPPEDSLMSLPLEESPLLPPEEPQLCPRSEPHLSRPREEPHLSRPREEPH
LSPQAEHPHLSPQPEEPCLCAVPPEEPHLSPQAEHPHLSPQPEELHLSPQTEEPHLSPVPE
EPCLSPQPEESHLSPQSEEPCLSPRPEESHLSPELEKPPLSRPEKPPPEPGQCPAPEEL
PLFPPPGEPSPSPLLGEPALSEPGEPPLSPLPEELPLSPSGEPSPSPLQMPDPPLPPPLS
PIITAAAPPALSPLGELEYFPFAGKSDSDPESPLAAPILETPISPPEANCTDPEVPVPMI
LPPSPGSPVGPASPILMELPQPQCSPLQHSPLVQNSPPSQCSPPALPLSVPSPLSPIGK
VVGVSDEAELHEMETEKVSEPECPALEPSATSPLSPMGDLSCPAPSPAPALDDFSGLGE
DTAPLDGIDAPGSQPEPGQTPGSLASELKGSPLLDPEELAPVTPMEVYPECKQTAGRS
PCEEQEPRAPVAPTPTLIKSDIVNEISNLSQGDASASFPGSEPLLGSPDPEGGSLSM
ELGVSTDVSPARDEGSLRLCTDSLPEPDDSLCDAGTAISGGKAEGEKRRRSSPARSRI
KQGRSSSFGRRRRPRGGAGHGRGRGRALRKSTASSIETLVVADIDSSPSKEEEEEDDDTM
QNTVVLFESNTDKFVLMQDMCVVCGSFGRGAEGHLLACSQCSQCYHPYCVNSKITKVMLLK
GWRCEVICVEVCGQASDPSRLLLCDDCDISYHTYCLDPPLLTVPKGGWKCKWCVSCMQC
GAASPGFHCEWQNSYTHCGPCASLVTCPICHAPYVEEDLLIQCRHCERWMHAGCESLFTE
DDVDHAPDEGFDVSCQPYVVKPVAPVAPPELVPMKVKEPEPQYFRFEGVWLTETGMALL
RNLTMSPHLKRRQRRGRGLPLGEAGLEGSEPSDALGPDDKKDGLDLDDELLKGEVVEHM
ECBIKLEGPVSPDVEPGKEETEESKRRKRKPYRPGIGFMVRQRKSHTRTKKGPAQAQAEV
LSGDGQPDVEIPADLPAEGAVEQSLAEGDEKKKQRRRGRKRSKLEGMPAYLQEAFFGKE
LLDLSRKALFAVGVRPSFGLGTPKAKGDGGERKELPTSQKGDDGPDIADEESRGLEGK
ADTPGPEDGGVKASPVPSDPEKPGTPGEGMLSSDLDRISTEELPKMESKDLQQLFKDVLG

SEREQHLGCGTPGLEGSRTPLQRPFLOQGLPLGNLPSSSPMDSYPGLCQSPFLDSRERG
 FFSPEPEPDSPTWGSSTPSTPTTTEGEGDGLSYNQSRSLQRWEKDEELGQLSTISP
 VLYANINFPNLKQDYPDWSSRCKQIMKLWRKVPAAADKAPYLQKAKDNRAAHRINKVQKQA
 ESQINKQTKVGDIAKRTDRPALHLRI PPQPGALGSPPPAAAPTIFIGSPPTPAGLSTAD
 GFLKPPAGSVPGPDSPELFLKLPPQVPAQAPSQDPFGLAPAYLEPRFPTAPPTYPPYP
 SPTGAPAQPPMLGASSRPGAGQPGFHTTTPGTPRHQSTPDPFLKPRCPSLDNLAVPES
 PGVGGGKASEPLLSPPPPGESRKALEVKKEELGASSPSYGPNNLGFVDSPPSGTHLGGLE
 LKTPDVFKAPLTPRASQVEPQSPGLGLRPQEPPPAALAPSPPSHPDIFRPGSYTDPYAQ
 PPLTPRPQPPPPESCCALPPRSLPSDPFSRVVSPSQSSSSQSPLTTPRPLSAEAFCPSPV
 TPRFQSPDPYSRPPSRPQSRDPFAPLHKPPRPQPPEVAFKAGSLAHTSLGAGGFPALPA
 GPAGELHAKVPSGQPPNFVRSPTGAFVGTTPSPMRFTFPQAVGEPSSLKPPVPQPGLP
 GINSHFPGPTLKGKQSTNYTVATGNFHPSGSPLGPSSGSTGESYGLSPLRPPSVLPPPA
 PDGSLPYLSHGASQSRGITSPEKREDPGTGMGSSLATAELPGTQDPGMSGLSQTELEKQ
 RQRQLRELLIRQQIQRNTRLQEKETAAAAAGAVPPGSGWGAEPSSPAFEQLSRGQTPFA
 GTQDKSSSLVGLPPSKLSGPI LGPGSFPSDDRLSRPPPPATPSSMDVNSRQLVGGSQAFYQ
 RPYPGSLPLQQQQQQLWQQQQAATAATSMRFAMSAFPPSTPGPELGRQALGSLAGISTR
 LPPGPEPVPGPAGPAQFIELRHNQKGLGPGGTFFPGQGPQRPRFYVSEDPHRLAPEG
 LRGLAVSGLPPQKPSAPPAPELNNSLHPTHTKGPPLTGLLELVNRPPSSTELGRNPPLA
 LEAGKLPCEDPELDDDFDAHKALEDDEELAHLLGLGVDVAKGDELGTLENLETNDPHLDD
 LLNGDEFDLLAYTDPELDTGDKKDI FNEHLRLVESANEEAEREALLRGVEPGPLGPEER
 PPAADASEPRLASVLPVEKPKVEEGGRHPSPCQFTIATPKVEPAPAANSLGLGLKPGQSM
 MGSRDTRMGTGPFSSSGHTAEKASFGATGGPPAHLTPSPPLSGPGSSLEKFELESGAL
 TLPGGPAASGDELDMESLVLASELPLLIEDLLEHEKKELQKQQLSAQLQPAQQQQQQQ
 QQHSLLPAPGPAQAMSLPHEGSSPSLAGSQQLSLGLAVARQPLPQPLMPTQPPAHALQ
 QRLAPSMAMVSNQGHMLSGQHGGQAGLVPPQSSQPVLSQKPMGTMPSPMCMKPPQLAMQ
 QLANSFFPDTDLDKFAAEDI IGPIAKAKMVALKGIKKVMAQGSIGVAPGMNRQVSLLAQ
 RLSSGGSSDLQNHVAAGSGQERSAGDPSQPRPNPTFAQGVINEADQRQYEEWLFTQQL
 LQMQLKVLLEEIGVHRKSRKALCAKQRTAKKAGREFPEADAELKLVTQQSKIQKQLDQ
 VRKQKQKHTNLMAEYRNKQQQQQQQQQQQHSVAVLALSPSQSPRLTLKPLGQLLPGHG
 LQPPQGPVGGQAGGLRLTPGGMALPGQPGGFLNTALAQQQQQQHSAGSLAGPSGGFF
 PGNLALRSLGPDRLQERQLQLQQQRMQLAQKLQQQQQQQQQHLGQVAIQQQQQG
 PGVQTNQALGPKPQGLMPPSSHQGLLVQQLSPQPPQGPQGMGLGPAQVAVLQQQHPGALGP
 QGPHRQVLMQTSRVLSSPQLAQQQQGLMGHRLVTAQQQQQQQQHQQQGSMAGLSHLQQL
 MSHSGQPKLSAQPMGSLQQLQQQQQLQQQQQLQQQQQLQQQQQLQQQQQLQQQQQLQQ
 QQQQQQLQQQQQLQQQQQQQQQQFQQQQQQQQMGLLNQSRLLSPQQQQQQQVALGPG
 MPAKPLQHFSSPGALGPTLLLTGKEQNTVDPAVSSATEGSPSTHQGGPLAIGTTPESMAT
 EPGEVKPSLSGDSQLLLVQPPQPPQSSQLQPPRLRPGQQQQQVSLHTAGGGSHGQLG
 SGSSSEASSVPHLLAQPSVSLGDQPGSMTQNLGPPQPMLEPMPQNNTPGPPKPGFVLQ
 SGQGLPGVIMPTVQQLRAQLQGVLAKNPQLRHLSPQQQQQLQALLMQRLQSSQAVRQT
 PPYQEPGTQTSPLQGLLGCQPLGQFPGPQTGPIQLGAGPRPQGPRLPAPPALSTGP
 VLGVPVHTPPPPSPQEPKPSQLPSPSSQLPTEAQLPPTHGTPKPKQGPTEPPPRVSP
 AAAQLADTLFSKGLGPDWPPDNLAETQKPEQSSLVPGHLDQVNGQVVEASQLSIKQEP
 EEPCLGAQSVKREANGEP I GAPGTSNHLHLAGPRSEAGHLLQKLLRAKNVQLSTGQGS
 EGLRAEINGHIDSKLAGLEQKLQGTSPSNKEDAAARKPLTPKPKRVQKASDRLVSSRKKLR
 KEDGVRASEALLQKQELSLPLTEPAITANFSLFAPFGSGCPVNGQSQRGAFGSGAL
 PTGPDYYSQLLTKNNLSNPPTPPSSLPPTPPPSVQKMNVTGPSEELGEHPKDAASARD
 SERALRDTSEVKSLDLLAALPTPHNQTEDVRMESEDESDSPDSIVPASSPESILGEEAP
 RFPHLGSGRWEQEDRALSPVIPLIPRDSIPVFPDTKPYGALGLEVPGLPVTTWEKGKGS
 EVSVMLTVSAAADKNLNGVMVAEALLSMKIPNSYEVLPESPARGGTEPKKGEAEGPGG
 KEKGLEKSPDTGPDWLKQFDAVLGYTLKRQLDILSLKQESPAPEPTQHRYTYNVSN
 LDVRQLSAPPEEPSPPPSPLASPASPTEPLVELPTEPLAEPVPSPLPLASSPESAR
 PKPRARPPEEGEDTRPPRLKWKGVWRKRRRLRGAMLELFGVNSLEVKFRTRSENGVLIHI
 QESSNYTTVIKNGKVYFTSDAGIAGKVERNIPEVYVADGHWHFTLIGKNGTATVLSVDR
 IYNRDI IHPTQDFGGLDVLITSLGGIPPNQAHRAQTAGFDGCIASMWYGGESLPFSGKH
 SLASISKTDPSVKIGCRGPNICASNPCWGDLLCINQWYAYRCVPPGDCASHPCQNGGSC
 PGLHSGFTCSPPDSHTGRTCEMVVACLGVLCPQKVKCAGSPAGHVCVLSQGPPEISLPL
 WAVPAIVGSCATVLALLVLSLILCNQCRGKKAKNPKEEKKPEKKKKGSENVAFDDPDNI
 PPYGDDMTVRKQPEGNPKPDI IERENPYLIYDETDIPHNSETIPSAPLASPEQIEHYDI
 DNASSIAPSDADI IQHYKQFRSHTPKFSIQRHSPLGFARQSPMPLGASSLTYPYQSYGQGL
 RNTSSLSHSACPTPNPLSRHSPAPFSKSSTFYRNSPARELHLPIRDGNTLEMHGDTCPGI
 FNYATRLGRRSKSPQAMASHGSRPGSRKLQPIGQIPLLESSPPVGLSIEEVERLNTPRPN
 PSICADHGRSSSEEDCRPLSRTRNPADGIPAPESSSDSDSHESFTCSEMEYDREKPMV
 YTSRMPKLSQVNESDADDEDNYGALRKPRRYHGRRAEGGPVGTQAAAPGTADNTLPMKLG
 QQAGTFNWDNLLNWPGFGHYVDVFKDLASLPEKAAANEKGAGTTKPVPKDGEAEQYV

A search of sequence databases reveals that the NOV47 amino acid sequence has 4776 of 4796 amino acid residues (99%) identical to, and 4779 of 4796 amino acid residues (99%) similar to, the 4957 amino acid residue ptrn:SPTREMBL-ACC:O14687 protein from Homo sapiens (Human) (ALR). Public amino acid databases include the GenBank databases,

5 SwissProt, PDB and PIR.

NOV47 is expressed in at least brain. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

10 The disclosed NOV47 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 47C.

Table 47C. BLAST results for NOV47					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 7512280 pir T03455	ALR protein - human	4957	4384/4427 (99%)	4385/4427 (99%)	0.0
gi 18601907 ref XP_028760.2 (XM_028760)	myeloid/lymphoid or mixed-lineage leukemia 2 [Homo sapiens]	3492	2930/2944 (99%)	2933/2944 (99%)	0.0
gi 3540281 gb AAC34383.1 (AF056116)	All-1 related protein [Takifugu rubripes]	4823	648/1458 (44%)	822/1458 (55%)	0.0
gi 13640139 ref XP_017017.1 (XM_017017)	protein FLJ23056 [Homo sapiens]	317	315/317 (99%)	315/317 (99%)	e-149
gi 10864041 ref NP_067053.1 (NM_021230)	myeloid/lymphoid or mixed-lineage leukemia 3; ALR-like protein [Homo sapiens]	4025	340/802 (42%)	467/802 (57%)	e-135

15 Tables 47D-H list the domain descriptions from DOMAIN analysis results against NOV47. This indicates that the NOV47 sequence has properties similar to those of other proteins known to contain this domain.

Table 47D. Domain Analysis of NOV47	
gnl Smart smart00282 , LamG, Laminin G domain	
CD-Length = 135 residues, 98.5% aligned	
Score = 85.1 bits (209), Expect = 9e-17	

Table 47E. Domain Analysis of NOV47

gnl|Pfam|pfam00628, PHD, PHD-finger. PHD folds into an interleaved type of Zn-finger chelating 2 Zn ions in a similar manner to that of the RING and FYVE domains.

CD-Length = 49 residues, 98.0% aligned

Score = 71.2 bits (173), Expect = 1e-12

Table 47F. Domain Analysis of NOV47

gnl|Pfam|pfam00769, ERM, Ezrin/radixin/moesin family. This family of proteins contain a band 4.1 domain (pfam00373), at their amino terminus. This family represents the rest of these proteins.

CD-Length = 365 residues, 23.0% aligned

Score = 52.8 bits (125), Expect = 5e-07

Table 47G. Domain Analysis of NOV47

gnl|Pfam|pfam00529, HlyD, HlyD family secretion protein.

CD-Length = 310 residues, 53.9% aligned

Score = 53.9 bits (128), Expect = 2e-07

Table 47H. Domain Analysis of NOV47

gnl|Pfam|pfam02166, Androgen_recep, Androgen receptor.

CD-Length = 456 residues, 18.6% aligned

Score = 52.8 bits (125), Expect = 5e-07

5 The ALL-1 gene is involved in human acute leukemia through chromosome
translocations or internal rearrangements. ALL-1 is the human homologue of Drosophila
trithorax. The latter is a member of the trithorax group (trx-G) genes which together with the
Polycomb group (Pc-G) genes act as positive and negative regulators, respectively, to
determine the body structure of Drosophila. ALR encodes a gigantic 5262 amino acid long
10 protein containing a SET domain, five PHD fingers, potential zinc fingers, and a very long run
of glutamines interrupted by hydrophobic residues, mostly leucine. The SET motif, PDH
fingers, zinc fingers and two other regions are most similar to domains of ALL-1 and TRX.
The first two motifs are also found in other trx-G and Pc-G proteins. The ALR gene was
mapped to chromosome band 12q12-13, adjacent to the VDR gene. This region is involved in
15 duplications and translocations associated with cancer.

The human ALL-1/MLL/HRX gene on chromosome 11q23 is the site of many locally
clustered chromosomal alterations associated with several types of acute leukemias, including

deletions, partial duplications and reciprocal translocations. Structurally variant proteins derived from an altered ALL-1 gene presumably make essential contributions to the malignant transformation of hematopoietic progenitor cells.

Many haematologic malignancies carry characteristic chromosomal translocations, which are thought to play an important role in the pathogenesis of these tumours. The t(8; 14) translocation in Burkitt's lymphoma was one of the first characterized at the molecular level. In this translocation the c-myc oncogene at chromosome 8q24 becomes deregulated by enhancer elements of the immunoglobulin heavy chain locus at chromosome 14q32 leading to a very aggressive B cell malignancy. Chromosomal translocations involving the MLL gene occur in about 80% of infant leukemias.

The disclosed NOV47 nucleic acid of the invention encoding a ALR-like protein includes the nucleic acid whose sequence is provided in Table 47A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 47A while still encoding a protein that maintains its ALR-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 1 percent of the bases may be so changed.

The disclosed NOV47 protein of the invention includes the ALR-like protein whose sequence is provided in Table 47B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 47B while still encoding a protein that maintains its ALR-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 1 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this ALR-like protein (NOV47) may function as a member of a "ALR family". Therefore, the NOV47 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential
5 therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

10 The NOV47 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the ALR-like protein (NOV47) may be useful in gene therapy, and the ALR-like protein (NOV47) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the
15 compositions of the present invention will have efficacy for treatment of patients suffering from Osteoporosis, involutional; Rickets, vitamin D-resistant; Fibrosis of extraocular muscles, congenital, 1; Achalasia-addisonianism-alacrimia syndrome; Cataract, polymorphic and lamellar; acute leukemias, cancers, or other pathologies or conditions. The NOV47 nucleic acid encoding the ALR-like protein of the invention, or fragments thereof, may further be
20 useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV47 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV47 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods
25 known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV47 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various
30 disorders.

NOV48

A disclosed NOV48 nucleic acid of 1988 nucleotides (also referred to as CG57713-01) encoding a sodium/bile acid transporter-like protein is shown in Table 48A. The start and stop codons are in bold letters.

5

Table 48A. NOV48 nucleotide sequence (SEQ ID NO:113).
<p>GGCGGCGACTGCGGCGACCGCGGGACGCGGAGAGGCACGCGCGGGAGGGGACCGGAATC CGCAGCTCCGGCCGCGCCATGGACGGCAACGACAACGTGACCTGCTCTTCGCCCTCTG CTGCGGGACAATACACCTGGCGCCCAATGCCAGCAGCCTGGGCCCGGGCACGGACCTC GCCCTCGCCCCCTGCCTCCAGCGCGCGGCCCGGCCCTGGGCTCAGCCTCGGGCCGGGTCCG AGCTTCGGCTTCAGCCCCGGCCCCACTCCGACCCCGGAGCCACGACCAGCGGCTCGCG GGCGGCGCGGGCAGCCACGGCCCTTCCCCGTTCCCTCGGCCCTGGGCGCCCCACGCGCTC CCGTTCTGGGACACGCCGCTGAACCACGGGCTGAACGTGTTCTGGGCGCCGCTGTGC ATCACCATGCTGGGCTGGGCTGCACGGTGGACGTGAACCACTTCGGGGCGCACGTCCGT CGGCCCGTGGGCGCGCTGCTGGCAGCGCTCTGCCAGTTCGGCTCCTGCCGCTGTGGCC TTCCTGCTGGCCCTCGCCTTCAAGCTGGACGAGGTGGCCCGCTGGCGGTGCTCCTGTGT GGCTGCTGTCCCGGGGCAATCTCTCAATCTTATGTCCCTGCTGGTTGACGGCGACATG AACCTCAGCATCATCATGACCATCTCTCCACGCTTCTGGCCCTCGTCTTGATGCCCTG TGCTGTGGATCTACAGCTGGGCTTGATCAACACCCCTATCGTGCAGTTACTACCCCTA GGGACCGTGACCTGACTCTCTGCAGCACTCTCATACCTATCGGGTGGGCGTCTTCATT CGCTACAAATACAGCCGGGTGGCTGACTACATGTGAAGGTAAGGCCCGTTTCCCTGTGG TCTCTGCTAGTGACTCTGGTGGTCCCTTTTATAATGACCGGCACTATGTTAGGACCTGAA CTGCTGGCAAGTATCCCTGCAGCTGTTTATGTGATAGCAATTTTATGCCTTTGGCAGGC TACGCTTCAGGTTATGGTTAGCTACTCTCTCCATCTTCCACCACTGCAAGAGGACT GTATGTCTGGAACAGGTAGTCAAGATGTGCAGCTCTGTACAGCCATTCTAAACTGGCC TTTCCACCGCAATTCATAGGAAGCATGTACATGTTTCTTTGCTGTATGCACTTTTCCAG TCTGCAGAAGCGGGGATTTTGTGTTTAAATCTATAAAATGTATGGAAGTGAAATGTTGCAC AAGCGAGATCCTCTAGATGAAGATGAAGATACAGATATTTCTTATAAAAACTAAAGAA GAGGAAATGGCAGACACTTCTATGGCAGTGAAGCAGAAAATATAATAATGATGGAA ACCGCTCAGACTTCTCTCAATATGGAGATACACAGGAGCTTCTATCTTGCTGAAATAT TGCTTCATATTATAGCTGTGGTAGTGACATGGTTAACATAAAAGATAACACTGGTTC ACATCATACATGTAACAATTCTGATCTTTTAAAGGTTCACTGGTGTATTAACCAAACGTT GTCACAAATTACAAATCAATGCTGTAATATAATTGCACCTGGAATGGCTAACGTGAAGC CTGAATTAAATGTGGTTTTFAGTTTTTACCATCACCAATTTCTATGACTGTTGCAATAC AGAATCTATTAGAAAACAGGGTCTTGGAATGTAGAATTTTGGCGCACTATGAGGAAAAA CAAGCTATCTTTGTAAAGCATAATTGAGTTTAAATGTAATTTGTTGTAATAAAAAAAGTGTG CTTGCTCTACTTAAATTCCTCACAATGTTGAATTTTGACCTGTATTGAGAAGAAATCCA AAACAGGTCAGTTAAATAAGGAAATATAGTATTTGTCAAACAGTATCAGAGAAAAGTTA CATTAATGATTTGATTACTTGATCTGGTATCTACTTATTAATGAATAATCAACATTTTT CTAGTGAA</p>

In a search of public sequence databases, the NOV48 nucleic acid sequence, located on chromosome 4 has 250 of 395 bases (63%) identical to a gb:GENBANK-ID:HUMNTCP|acc:L21893.1 mRNA from Homo sapiens (Human Na/taurocholate cotransporting polypeptide mRNA, complete cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

10

The disclosed NOV48 polypeptide (SEQ ID NO:114) encoded by SEQ ID NO:113 has 440 amino acid residues and is presented in Table 48B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV48 has no signal peptide and is likely to be localized in the plasma membrane with a certainty of 0.6000.

15

Table 48B. Encoded NOV48 protein sequence (SEQ ID NO:114).

MDGNDNVTLFLFAPLLRDNYTLAPNASSLGPGTDLALAPASSAGPGPGLSLGPGPSFGFSP
 GPTPTPEPTTSLAGGAASHGPSFPFPRPWAPHALPFWDTPLNHGLNVFVGAALCITMLGL
 GCTVDVNHFGAHVRRPVGALLAALCQFGLPLLAFLALAFKLDEVAAVAVLLCGCCPGG
 NLSNLSLLVDGDMNLSIIMTISSTLLALVLMPLCLWIYSWAWINTPIVQLPLGTVTLT
 LCSTLIPIGLGVFIRYKYSRVADYIVKVRPVSLSLLVTLVVLFIIMTGTMLGPELLASIP
 AAVYVIAIFMPLAGYASGYGLATLFLHLPNCKRRTVCLETGSQNVQLCTAILKLAFPPQFI
 GSMYMFLLYALFQSAEAGIFVLIYKMYGSEMLHKRDPLDEDEDTDISYKKLKEEEMADT
 SYGTVKAENIIMMETAQTSL

A search of sequence databases reveals that the NOV48 amino acid sequence has 126 of 325 amino acid residues (38%) identical to, and 193 of 325 amino acid residues (59%) similar to, the 362 amino acid residue ptnr:SWISSPROT-ACC:O08705 protein from Mus musculus (Mouse) (SODIUM/BILE ACID COTRANSPORTER (NA(+)/BILE ACID COTRANSPORTER) (NA(+)/TAUROCHOLATE TRANSPORT PROTEIN) (SODIUM/TAUROCHOLATE COTRANSPORTING POLYPEPTIDE)). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV48 is expressed in at least Liver. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV 48 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 48C.

Table 48C. BLAST results for NOV48

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 15082287 gb AAH12048.1 AAH12048 (BC012048)	protein for IMAGE:3502817) [Homo sapiens]	467	437/440 (99%)	437/440 (99%)	0.0
gi 17512162 gb AAH19066.1 AAH19066 (BC019066)	protein for MGC:29802) [Homo sapiens]	432	437/440 (99%)	437/440 (99%)	0.0
gi 15294592 ref XP_053248.1 (XM_053248)	protein XP_053248 [Homo sapiens]	435	358/445 (80%)	365/445 (81%)	e-66
gi 3980315 emb CAA10360.1 (AJ131361)	hepatic sodium- dependent bile acid transporter [Oryctolagus cuniculus]	348	125/334 (37%)	192/334 (57%)	8e-54

cDNA encoding the rat liver bile acid uptake system has been isolated by expression cloning in *Xenopus laevis* oocytes. The cloned transporter is strictly sodium-dependent and can be inhibited by various non-bile-acid organic compounds. Sequence analysis of the cDNA revealed an open reading frame of 1086 nucleotides coding for a protein of 362 amino acids (calculated molecular mass 39 kDa) with five possible N-linked glycosylation sites and seven putative transmembrane domains. Translation experiments in vitro and in oocytes indicate that the transporter is indeed glycosylated and that its polypeptide backbone has an apparent molecular mass of 33-35 kDa. Northern blot analysis with the cloned probe revealed crossreactivity with mRNA species from rat kidney and intestine as well as from liver tissues of mouse, guinea pig, rabbit, and man. PMID: 1961729

Using expression cloning in *Xenopus laevis* oocytes, a cDNA encoding a rat liver organic anion-transporting polypeptide (oatp) was isolated. The cloned oatp mediated Na(+)-independent uptake of sulfobromophthalein (BSP) which was Cl(-)-dependent in the presence of bovine serum albumin (BSA) at low BSP concentrations (e.g., 2 microM). Addition of increasing amounts of BSA had no effects on the maximal velocity of initial BSP uptake, but it increased the Km value from 1.5 microM (no BSA) to 24 microM (BSA/BSP molar ratio, 3.7) and 35 microM (BSA/BSP ratio, 18.4). In addition to BSP, the cloned oatp also mediated Na(+)-independent uptake of conjugated (taurocholate) and unconjugated (cholate) bile acids. Sequence analysis of the cDNA revealed an open reading frame of 2010 nucleotides coding for a protein of 670 amino acids (calculated molecular mass, 74 kDa) with four possible N-linked glycosylation sites and 10 putative transmembrane domains. Translation experiments in vitro indicated that the transporter was indeed glycosylated and that its polypeptide backbone had an apparent molecular mass of 59 kDa. Northern blot analysis with the cloned probe revealed crossreactivity with several mRNA species from rat liver, kidney, brain, lung, skeletal muscle, and proximal colon as well as from liver tissues of mouse and rabbit, but not of skate (*Raja erinacea*) and human. PMID: 8278353

Active uptake of bile acids from the lumen of the small intestine is mediated by an ileal Na(+)-dependent bile acid transport system. To identify components of this transport system, an expression cloning strategy was employed to isolate a hamster ileal cDNA that exhibits bile acid transport activity. By Northern blot analysis, mRNA for the cloned transporter was readily detected in ileum and kidney but was absent from liver and proximal small intestine. The transporter cDNA encoded a 348-amino acid protein with seven potential transmembrane domains and three possible N-linked glycosylation sites. The amino acid sequence was 35% identical and 63% similar to the rat liver Na+/bile acid cotransporter. After transfection into

COS cells, the hamster cDNA transported taurocholate in a strict Na(+)-dependent fashion with an apparent Km of 33 microM. This taurocholate transport was inhibited by various bile acids but not by taurine or other organic anions. The Na+ dependence, saturability, and bile acid specificity of transport as well as the tissue specificity of mRNA expression strongly argue that the transporter cDNA characterized in this study is the Na+/bile acid cotransporter described previously in ileum. PMID: 8288599

Uptake of long-chain and aromatic neutral amino acids into cells is known to be catalyzed by the Na(+)-independent system L transporter, which is ubiquitous in animal cells and tissues.. The 2.3-kilobase cDNA codes for a protein of 683 amino acids. The transporter has four putative membrane-spanning domains and bears no sequence or structural homology to any known animal or bacterial transporter. When transcribed and expressed in *Xenopus* oocytes, the transporter exhibits many, but not all, of the characteristics of L-system transporters, suggesting that this represents one of several related L-system transporters. PMID: 1729674

The phylogenic and ontogenic expression of mRNA for the Na+/bile acid cotransporter was determined by Northern analysis utilizing a full-length cDNA probe recently cloned from rat liver. mRNA was detected in several mammalian species, including rat, mouse, and man, but could not be found in livers from nonmammalian species, including chicken, turtle, frog, and small skate. When expression of the bile acid transporter in developing rat liver was studied, mRNA was detected between 18 and 21 days of gestation, at the time when Na(+)-dependent bile acid transport is first detected. Two hepatoma cell lines (HTC and HepG2), the latter of which is known to have lost the Na+/bile acid cotransport system, also did not express mRNA for this transporter. Finally, when mRNA from the lower vertebrate (the small skate) was injected into *Xenopus* oocytes, only a sodium-independent, chloride-dependent transport system for bile acids was expressed, confirming the integrity of the mRNA and consistent with prior functional studies of bile acid transport in this species. These findings establish that the Na+/bile acid cotransport mRNA is first transcribed in mammalian species, a process that is recapitulated late during mammalian fetal development in rat liver, and that this mRNA is lost in dedifferentiated hepatocytes. In contrast, the mRNA for a multispecific Na+/independent organic anion transport system is transcribed earlier in vertebrate evolution. PMID: 8421672

The disclosed NOV48 nucleic acid of the invention encoding a sodium/bile acid transporter-like protein includes the nucleic acid whose sequence is provided in Table 48A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 48A while still encoding a

5
10

15

 $(F_{ab})_2,$

20

25

30

efficacy for treatment of patients suffering from cancer, trauma, regeneration (in vitro and in vivo), viral/bacterial/parasitic infections, Von Hippel-Lindau (VHL) syndrome, Cirrhosis, Transplantation, or other pathologies or conditions. The NOV48 nucleic acid encoding the sodium/bile acid transporter-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV48 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV48 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV48 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV49

A disclosed NOV49 nucleic acid of 2313 nucleotides (also referred to as CG57721-01) encoding a prestin-like protein is shown in Table 49A. The start and stop codons are in bold letters.

Table 49A. NOV49 nucleotide sequence (SEQ ID NO:115).

CCATTGACTGCAGGAAGGTTGGCCAGCAGAGCAAATGCCATGCCTGCAGAGGACAACGAG
ACTGAAGCTCAGCAGCATGACCGCTACGTGGTAGACAGAGCCGCATACCTCCCTTACCCTC
TTCGACGATGAGTTTGAAGAAGGACCGGACATACCCAGTGGGAGAGAACTTCGCAAT
GCCTTCAGATGTTCTCAGCCAAGATCAAAGCTGTGGTGTGTTGGGCTGCTGCCTGTGCTC
TCCTGGCTCCCAAGTACAAGATTAAAGACTACATCATTCCTGACCTGCTCGGTGGACTC
AGCGGGGATCCATCCAGGTCCCAAGGTATGGCATTGTCTGTCTGGCCAACCTTCCT
GCAGTCAATGGCCTCTACTCCTCTCTTCCCCCTCTGACCTACTTCTCCTGGGGGT
GTTCAACAGATGGTGCCAGGTACCTTTGCCGTTATCAGCATCCTGGTGGGTAACATCTGT
CTGCAGCTGGCCCCAGAGTCGAAATCCAGGTCTTCAACAATGCCACCAATGAGAGCTAT
GTGGACACAGCAGCCATGGAGGCTGAGAGGCTGCACGTGTGAGTACGCTAGCCTGCCTC
ACCGCCATCATCCAGATGGGTCTGGGCTTCATGCAGTTTGGCTTTGTGGCCATCTACCTC
TCCGAGTCCTTCATCCGGGGCTTCATGACGGCCCGCGCTGCAGATCCTGATTTCCGCTG
CTCAAGTACATCTTCGACTGACCATCCCCCTCTACACAGGCCAGGGTCCATCGTCTTT
GTGAGTCTGGGGATGTGCAAAAACCTCCCCACACCAACATCGCCTCGCTCATCTCGCT
CTCATCAGCGGTGCCTTCCTGGTGTGGTGAAGGAGCTCAATGCTCGTACATGCACAAG
ATTGCTTTCCCCATCCCTACAGAGATGATTGTGGTAAGGACCTTGTTCAAGCTGGGTGT
AAGATGCCCAAAAAGTATCACATGAGATCGTGGGAGAAATCCAACCTCGGCAGGTTCCTC
ACCCCGGTGTCGCTGTGGTCTCACAGTGAAGGACATGATAGGCACAGCCTTCTCCCTA
GCCATCGTGAAGTACGTCATCAACCTGGCTATGGGCCGACCTGGCCAACAAGCAGGC
TACGACGTGGATTGCAACCAGGAGATGATCGCTCTCGGCTGCAGCAACTTCTTTGGCTCC
TTCTTTAAATTCATGTCATTTGCTGTGCGCTTCTGTCACTCTGGCTGTGGATGGAGCT
GGAGGAAAATCCAGGTGAGCCTTGTTCTAGGGGAGTTGTCTGAGCTCCCTTCTTACTC
ACCACGGGGTTTGCTTAAGAGTACTCAGGTGTCTCTGTGCTAGGAGCCCTGATCGCT
GTCAATCTCAAGAACTCCCTCAAGCAACTACCGACCCCTACTACCTGTGGAGGAAGAGC

AAGCTGGACCACTGTCATCTGGGTAGTGAGCTTCCTCTCCTCTCTCTCCTCAGCCTGCC
TATGGTGTGGCAGTGGGTGTGCGCTTCTCCGTCTGGTCTGGTCTTCCAGACTCAGAGT
CGAAATGGCTATGCACTGGCCAGGTCACTGGACACTGACATTATGTGAATCCCAAGACC
TATAATAGGGTACAGGATATCCAGGGGATTAATATCATCAGTACTGCTCCCCTCTCTAC
TTTGCCAACTCAGAGATCTTCAGGCAAAAGGTCACTGCGCAAGGTAAGGCTCAGTCCCTGG
CGACCAGAGGCTCTGGACAGAGAGTGGCCGAAATGGAAGCAGAAGGGCGGTGGGACCC
AACAAACAACAGACCCCGGCTAACGGCACAGCGTGTCTATATCACCTTCAGCCCTGAC
AGCTCCTCACCTGCCAGAGTGAGCCACCAGCCTCCGCTGAGGCCCCCGGCGAGCCAGT
GACATGCTGGCCAGCGTCCACCCCTTCGTACCTCCACACCTCATCCTCGACATGAGT
GGAGTCAGCTTCGTGGACTTGATGGGCATCAAGGCCCTGGCCAAGCTGAGTCCACCTAT
GGGAAGATCGGCGTAAAGGTCTTCTTGGTGAACATCCATGCCAGGTGTACAATGACATT
AGCCATGGAGGCGTCTTTGAGGATGGGAGCCTAGGATGCAAGCAGTCTTTCCAGCATA
CATGACGCGAGTCTCTTTGCCAGCTGATTGAGTTACCTGGATTGAGGTCAGTGGCAATG
GCTGAAGTGGAGACGAGGTGGAAGTGGTTCAGGCCGGGGAATCACCACTTGAGTTTG
TACTAAAGACCCAGCCAGCCCTGTTTCTCTT

In a search of public sequence databases, the NOV49 nucleic acid sequence, located on chromosome 3 has 966 of 1618 bases (59%) identical to a gb:GENBANK-

ID:AF279265|acc:AF279265.1 mRNA from Homo sapiens (Homo sapiens putative anion

5 transporter 1 mRNA, complete cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV49 polypeptide (SEQ ID NO:116) encoded by SEQ ID NO:115 has 748 amino acid residues and is presented in Table 49B using the one-letter amino acid

10 code. Signal P, Psort and/or Hydropathy results predict that NOV49 has a signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.6000. The most likely cleavage site for a NOV49 peptide is between amino acids 8 and 9.

Table 49B. Encoded NOV49 protein sequence (SEQ ID NO:116).

MPAEDNETEAQQHDRVVDRAAYSLTLFDFEFKDRTPVGEKLRNFRCSAKIKAVV
FGLLPVLSWLPKYKIDYIIPDLLGGLSGGSIQVPQGMFAFALLANLPAVNGLYSSFFPLL
TYFFLGGVHQMPGTFAVISILVGNICQLAPESKFQVFNNATNESYVDTAAMEAERLHV
SATLACLTAIQMGLGFMQGFVAIYLSSEFIRGFMTAAGLQILISVLKYIFGLTIPSYT
GPGSIVFVSLGMCKNLPHTNIASLIFALISGAFVLVLKELNARYMHKIRFPPIPTMIVVR
TLFRAGCKMPKKYHMQIVGEIQLGRFPTPVSPVVSQWKDMIGTAFSLAIVSYVINLAMGR
TLANKHGYDSDSNQEMIALGCSNFFGSFFKIHVICCALSVTLAVDGAGGKSQVSLVLGEL
SELPFLLLTTGFALRVLRLCLSVLGALIAVNLKNSLKLQTLDPYYLWRKSKLDQCIWVVSFLS
SFFLSLPYGVAVGVAFSVLVVVFQTQSRNGYALAQVMDTDIYVNPPTYNRVQDIQGIKII
TYCSPLYFANSEIFRQKVIKVRVLSRPEALDREWPENGSRRAVGPNNNQTPANGTSVS
YITFSPDSSSPAQSEPPASAEAPGEPDMLASVPPFVTFHTLILDMSGVSFVDFLMGIKAL
AKLSSTYKGIGVKVFLVNIHAQVYNDISHGGVFEDGSLGCKHVFPSIHDAVLFAQLIQLP
GLRSLAMAEVETQVELVQAGGITHLSLY

A search of sequence databases reveals that the NOV49 amino acid sequence has 277
of 732 amino acid residues (37%) identical to, and 434 of 732 amino acid residues (59%)
15 similar to, the 744 amino acid residue ptrn:TREMBLNEW-ACC:CAC21555 protein from
Rattus norvegicus (Rat) (PRESTIN). Public amino acid databases include the GenBank
databases, SwissProt, PDB and PIR.

NOV49 is expressed in at least Parotid Salivary glands. Expression information was derived from the tissue sources of the sequences that were included in the derivation of the sequence of CG57721-01. The sequence is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: gb:GENBANK-

5 ID:AF279265|acc:AF279265.1) a closely related Homo sapiens putative anion transporter 1 mRNA, complete cds homolog in species Homo sapiens :kidney and pancreas. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

10 The disclosed NOV49 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 49C.

Table 49C. BLAST results for NOV49					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 16588681 gb AAL2 6867.1 AF314958.1 (AF314958)	anion transporter/excha nger-9 [Homo sapiens]	887	642/739 (86%)	657/739 (88%)	0.0
gi 16418413 ref NP 443166.1 (NM_052934)	solute carrier family 26, member 9 [Homo sapiens]	791	642/739 (86%)	657/739 (88%)	0.0
gi 15011891 ref NP 109652.2 (NM_030727)	prestin (motor protein) [Mus musculus]	744	281/748 (37%)	430/748 (56%)	e-129
gi 13540646 ref NP 110467.1 (NM_030840)	prestin [Rattus norvegicus]	744	278/746 (37%)	435/746 (58%)	e-128
gi 8050590 gb AAF71 715.1 AF230376.1 (AF230376)	prestin [Meriones unguiculatus]	744	271/738 (36%)	424/738 (56%)	e-125

15 Tables 49D-E list the domain descriptions from DOMAIN analysis results against NOV49. This indicates that the NOV49 sequence has properties similar to those of other proteins known to contain this domain.

Table 49D. Domain Analysis of NOV49
gnl Pfam pfam00916, Sulfate_transp, Sulfate transporter family.
Mutations in human Diastrophic Dysplasia Protein lead to several diseases.
CD-Length = 312 residues, 100.0% aligned
Score = 167 bits (423), Expect = 2e-42

Table 49E. Domain Analysis of NOV49

gnl|Pfam|pfam01740, STAS, STAS domain. The STAS (after Sulphate Transporter and AntiSigma factor antagonist) domain is found in the C terminal region of Sulphate transporters and bacterial antisigma factor antagonists. It has been suggested that this domain may have a general NTP binding function.

CD-Length = 106 residues, 62.3% aligned

Score = 57.0 bits (136), Expect = 4e-09

A second distinct family of anion transporters, in addition to the classical SLC4 (or AE) family, has recently been delineated. Members of the SLC26 family are structurally well conserved and can mediate the electroneutral exchange of Cl⁻ for HCO⁻(3) across the plasma membrane of mammalian cells like members of the SLC4 family. Three human transporter proteins have been functionally characterized: SLC26A2 (DTDST), SLC26A3 (CLD or DRA), and SLC26A4 (PDS) can transport with different specificities the chloride, iodine, bicarbonate, oxalate, and hydroxyl anions, whereas SLC26A5 (prestin) was suggested to act as the motor protein of the cochlear outer hair cell.

Electromotility, i.e., the ability of cochlear outer hair cells (OHCs) to contract and elongate at acoustic frequencies, is presumed to depend on the voltage-driven conformational changes of "motor" proteins present in the OHC lateral plasma membrane. Recently, two membrane proteins have been proposed as candidates for the OHC motor. A sugar transporter, GLUT-5, was proposed based on its localization in the OHCs and on the observation that sugar transport alters the voltage sensitivity of the OHC motor mechanism. Another candidate, "prestin," was identified from a subtracted OHC cDNA library and shown to impart voltage-driven shape changes to transfected cultured cells.

The disclosed NOV49 nucleic acid of the invention encoding a prestin-like protein includes the nucleic acid whose sequence is provided in Table 49A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 49A while still encoding a protein that maintains its prestin-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least

in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 41 percent of the bases may be so changed.

5 The disclosed NOV49 protein of the invention includes the prestin-like protein whose sequence is provided in Table 49B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 49B while still encoding a protein that maintains its prestin-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 63
10 percent of the residues may be so changed.

 The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

 The above defined information for this invention suggests that this prestin-like protein (NOV49) may function as a member of a "prestin family". Therefore, the NOV49 nucleic
15 acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug
20 targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

 The NOV49 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the prestin-like protein
25 (NOV49) may be useful in gene therapy, and the prestin-like protein (NOV49) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from Autoimmune disease, Renal artery stenosis, Interstitial nephritis, Glomerulonephritis, Polycystic kidney disease, Systemic lupus erythematosus, Renal tubular acidosis, IgA
30 nephropathy, Hypercalcaemia, Lesch-Nyhan syndrome, Diabetes, Von Hippel-Lindau (VHL) syndrome, Pancreatitis, Obesity, Xerostomia, cancer, trauma, regeneration (in vitro and in vivo), viral/bacterial/parasitic infections, or other pathologies or conditions. The NOV49 nucleic acid encoding the prestin-like protein of the invention, or fragments thereof, may

further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV49 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV49 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV49 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

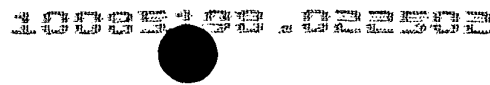
NOV50

A disclosed NOV50 nucleic acid of 1335 nucleotides (also referred to as CG57787-01) encoding a sulfate transporter-like protein is shown in Table 50A. The start and stop codons are in bold letters.

Table 50A. NOV50 nucleotide sequence (SEQ ID NO:117).

GCCTTCCTGTCGGGCTGCATCCAGCTGGCCATGGGGTCTGCGTTTGGGGTTCTGCTGGACTTCATTT
CCTACCCCGTCATTAAAGGCTTCACCTCTGCTGCTGCCGTACCATCGGCTTTGGACAGATCAAGAACCT
GCTGGGACTACAGAACATCCCCAGGCCGTTCTTCTGTCAGGTGTACCACACCTTCCTCAGGATTGCAGAG
ACCAGGGTAGGTGACGCCGTCCCTGGGGCTGGTCTGCATGTGCTGCTGCTGGTGCTGAAGCTGATGCGGG
ACCACGTGCCTCCCGTCCACCCGAGATGCCCCCTGGTGTGCGGCTCAGCCGTGGGCTGGTCTGGGCTGC
CAGACAGCTCGCAACGCCCTGGTGGTCTCCTTCGCAGCCCTGGTTGCGTACTCCTTCGAGGTGACTGGA
TACCAGCCTTTCATCCTAACAGGGGAGACAGCTGAGGGGCTCCCTCCAGTCCGGATCCCGCCCTTCTCAG
TGACCACAGCCAACGGGACGATCTCCTTCACCGAGATGGTGCAGGACATGGGAGCCGGGCTGGCCGTGGT
GCCCCCTGATGGGCCTCCTGGAGAGCATTGCGGTGGCCAAAGCCTTCGCATCTCAGAATAATTACCGCATC
GATGCCAACAGGAGCTGCTGGCCATCGGTCTCACCACATGTTGGGCTCCCTCGTCTCCTCCTACCCGG
TCACAGGCAGCTTTGGACGGACAGCCGTGAACGCTCAGTCGGGGGTGTGCACCCCGCGGGGGGCTGGT
GACGGGAGTGCTGGTGTCTCTGGACTACCTGACCTCACTGTTCTACTACATCCCCAAGTCTGCC
CTGGCTGCCGTATCATCATGAGCCGTGGCCCCGCTGTTTCGACACCAAGATCTTCAGGACGCTCTGGCGTG
TTAAGAGGCTGGACCTGCTGCCCCGTGCGTGACCTTCCTGCTGTGCTTCTGGGAGGTGCAGTACGGCAT
CCTGGCCCGGGCCCTGGTGTCTCTGCTCATGCTCCTGCACCTCTGCAGCCAGGCCTGAGACCAAGGTGTCA
GAGGGGCCGGTTCTGGTCTGCAGCCGGCCAGCGGCCTGTCTTCCCTGTCTCTGCCCCCACTCCCTG
CTGTTTACAGGACCCCAAGACCCTGTCCCGACGCTCTCCAGTCCACAAGGATGCAGGCATCTCTGAGTGGG
CTGGACCGTCTCTGTGGCCCTCAGCCAGTGGTGTGCTGCAGCAAGGGTGGTGGCTCCCCACATATCACTCC
TTCCCTGCCCTAAAGTCCGGTTCCTGTTTCTGGGGGTGATTATTTAGGGGAGCTAAGGGCCTGTGAGTC
CTAGT

In a search of public sequence databases, the NOV50 nucleic acid sequence has 585 of 993 bases (58%) identical to a gb:GENBANK-ID:AF297659|acc:AF297659.2 mRNA from Homo sapiens (Homo sapiens sulfate/anion transporter SAT-1 protein (SLC26A1) mRNA, complete cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.



The disclosed NOV50 polypeptide (SEQ ID NO:118) encoded by SEQ ID NO:117 has 384 amino acid residues and is presented in Table 50B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV50 has a signal peptide and is likely to be localized extracellularly with a certainty of 0.6000. The most likely cleavage site for a NOV50 peptide is between amino acids 20 and 21.

Table 50B. Encoded NOV50 protein sequence (SEQ ID NO:118).
MGVLR LGFLDFISYPVIKGFTSAAAVTIGFGQIKNLLGLQNI PRPF FLQVYHTFLRIAE TRVGDAVLGLVCM LLLVLKLMRDHVPVHPMP PGVRLSRGLVWAATTARNALVVSFAA LVAYSFEVTGYQPFI LTGETAEGLPVRIPPF SVTTANGTISFTEMVQDMGAGLAVVPLM GLLESIAVAKAFASQNNYRIDANQELLAIGLTNMLGSLVSSYPVTGSFGR TAVNAQSGVC TPAGGLVTGVLVLLSLDYLTSLFYIIPKSALA AVIIMAVAPLFDTKIFRTLWRVKRLDLL PLCVTFLLCFWEVQYGILAGALVSLMLLHSAARPETKVSEGPVLVLQPASGLSFPVLC P PLPAVQDPKTLSP TLLSSPQGCRHL

A search of sequence databases reveals that the NOV50 amino acid sequence has 146 of 339 amino acid residues (43%) identical to, and 210 of 339 amino acid residues (61%) similar to, the 595 amino acid residue ptnr:SPTREMBL-ACC:Q9V879 protein from *Drosophila melanogaster* (Fruit fly) (CG5002 PROTEIN). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV50 is expressed in at least Brain, Cerebral Medulla/Cerebral white matter, Coronary Artery, Epidermis, Frontal Lobe, Hippocampus, Hypothalamus, Kidney, Liver, Lung, Mammary gland/Breast, Oviduct/Uterine Tube/Fallopian tube, Pituitary Gland, Retina, Spinal Chord, Spleen, Substantia Nigra, Temporal Lobe, Testis, Umbilical Vein, Whole Organism. Expression information was derived from the tissue sources of the sequences that were included in the derivation of the sequence of CG57787_01. The sequence is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:AF297659|acc:AF297659.2) a closely related *Homo sapiens* sulfate/anion transporter SAT-1 protein (SLC26A1) mRNA, complete cds homolog in species *Homo sapiens* :liver. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV50 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 50C.

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
<u>gi 18485058 ref XP_080006.1 </u> (XM_080006)	CG5002 [Drosophila melanogaster]	595	121/341 (35%)	169/341 (49%)	1e-46
<u>gi 7301216 gb AAF56347.1 </u> (AE003749)	Esp gene product [Drosophila melanogaster]	654	124/368 (33%)	180/368 (48%)	7e-44
<u>gi 17738183 ref NP_524490.1 </u> (NM_079766)	Epidermal stripes and patches [Drosophila melanogaster]	623	127/370 (34%)	179/370 (48%)	3e-43
<u>gi 7301881 gb AAF56989.1 </u> (AE003772)	CG7912 gene product [Drosophila melanogaster]	573	120/323 (37%)	170/323 (52%)	4e-42
<u>gi 7300023 gb AAF55195.1 </u> (AE003708)	CG6125 gene product [Drosophila melanogaster]	640	100/315 (31%)	162/315 (50%)	3e-38

5

Table 50D. Domain Analysis of NOV50

gnl|Pfam|pfam00916, Sulfate_transp, Sulfate transporter family.

Mutations in human Diastrophic Dysplasia Protein lead to several diseases.

CD-Length = 312 residues, 95.2% aligned

Score = 163 bits (413), Expect = 1e-41

10

15

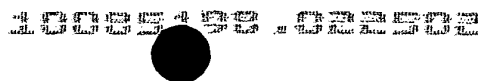
their differential sensitivity to the DIDS anion-exchanger inhibitor. They reported the molecular characterization of a DIDS-resistant sulfate transporter from human HEVECs, designated SUT1. SUT1 belongs to the family of sodium-coupled anion transporters and exhibits 40 to 50% amino acid identity with the rat renal sodium/sulfate cotransporter NaSi1, as well as with the human and rat sodium/dicarboxylate cotransporters NADC1/SDCT1 (604148) and NADC3/SDCT2.

The disclosed NOV50 nucleic acid of the invention encoding a sulfate transporter-like protein includes the nucleic acid whose sequence is provided in Table 50A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 50A while still encoding a protein that maintains its sulfate transporter-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 42 percent of the bases may be so changed.

The disclosed NOV50 protein of the invention includes the sulfate transporter-like protein whose sequence is provided in Table 50B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 50B while still encoding a protein that maintains its sulfate transporter-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 57 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this sulfate transporter-like protein (NOV50) may function as a member of a "sulfate transporter family". Therefore, the NOV50 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to:



protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

5 The NOV50 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the sulfate transporter-like protein (NOV50) may be useful in gene therapy, and the sulfate transporter-like protein (NOV50) may be useful when administered to a subject in need thereof. By way of
10 nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from diastrophic dysplasia and certain other skeletal dysplasias, and adenoma, familial chloride diarrhea, CNS disorders, brain disorders including epilepsy, eating disorders, schizophrenia, ADD; cancer; heart disease; inflammation and autoimmune disorders including Crohn's disease, IBD, allergies, rheumatoid and osteoarthritis,
15 inflammatory skin disorders, blood disorders; psoriasis colon cancer, leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, pancreatic disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer, or other pathologies or conditions. The NOV50 nucleic acid encoding the sulfate transporter -like protein of the
20 invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV50 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV50 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods
25 known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV50 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various
30 disorders.

NOV51

A disclosed NOV51 nucleic acid of 2079 nucleotides (also referred to as CG57785-01) encoding a sulfate transporter-like protein is shown in Table 51A. The start and stop codons are in bold letters.

5

Table 51A. NOV51 nucleotide sequence (SEQ ID NO:119).
<p>ATGGTTGTGGCTGTACAGATTTTAACTCCTCGGTGCACCAAGAGTACAGATTCCAAATTGCCTATATCT TAAAAACTTGTAAAGAAAGAAGCAATGGTGAATGTGAATCTGAACACCAGGGAGAGTTCAGAAAGGGAAT TCCCATCAGTTGGTACTACCTAATAATGGGCGTATTGGGTTTGGGCTTCATTGCCACTTACCTTCCGGAG TCTGCAATGAGTGCTTACCTGGCTGCTGTGGCACTTCATATCATGCTGTCCAGCTGACTTTCATCTTTG GGATTATGATTAGTTTCCATGCCGGTCCCATCTCCTTCTTCTATGACATAATTAATTACTGTGTAGCTCT CCCCAAAGCGAATTCACCAGCATTCTAGTATTTCTAACTGTGTTGTGCTCTGCGAATCAACAAATGT ATCAGAATTTCTTCAATCAGTATCCCATGAGTTTCCCATGGAATTAATTTCTGATTATTGGCTTCACTG TGATTGCAAAACAAGATAAGCATGGCCACAGAAACCAGCCAGACGCTTATGACATGATTCCCTTATAGCTT TCTGCTTCTGTAAACACCAGATTTTCAGCCTTCTTCCCAAGATAATTTTACAAGCCTTCTCCTTATCTTTG GTGAGCTCCTTTCTGCTCATATTTCTGGGCAAGAAGATTGCCAGTCTTACAATTACAGTGTCAATTCCA ACCAGGATTTAATAGCCATCGGCCTTTGCAATGTCGTCAAGTTCATTTTTCAGATCTGTGTGTTTACTGG TGCTATTGCTAGGACTATTATCCAGGATAAATCTGGAGGAAGACAACAGTTTGCATCTCTGGTAGGCGCA GGTGTGATGCTCCTGATGGTGAAGATGGGACACTTTTTCTACACACTGCCAAATGTTGATATGGTAA AGGTGCCTCTTAAAGAAGAAGAAATTTTCAGCTTGTTTAATTCAAGTGACACCAATCTACAAGGAGGAAA GATTTGCAGGTGTTTCTGCAACTGTGATGATCTGGAGCCGCTGCCAGGATTCTTTACACAGAGCGATTT GAAATAAACTGGATCCCGAAGCATCCTCCATTAACCTGATTCACTGCTCACATTTTGAGAGCATGAACA CAAGCCAAACTGCATCCGAAGACCAAGTGCCATACACAGTATCGTCCGTGCTCAGAAAAATCAAGGGCA ACAGTATGAGGAGGTGGAGGAAGTTTGGCTTCTTAATAACTCATCAAGAAACAGCTCACCAGGACTGCCT GATGTGGCGGAAGCCAGGGGAGGAGATCACTCATCCCTTACTCAGATGCGTCTCTACTGCCAGTGTCC ACACCATCATCCTGGATTCTCCATGGTACACTACGTGGATTACGGGGGTTAGTCGTATTAAGACAGAT ATGCAATGCCTTTCAAAACGCCAACATTTTGATACTCATTGCAGGGTGTCACTCTCCATAGTCAGGGCA TTTGAGAGGAATGATTTCTTTGACGCTGGCATCACCAGACCCAGCTGTTCTCAGCGTTCAGACGCCG TGCTGTTTGCCTTGTCAAGGAAGGTTCATAGGCTCCTCTGAGTTAAGCATCGATGAATCCGAGACAGTGAT ACGGGAAACCTACTCAGAAACAGACAGAATGACAATTCAGATATAAAATGAGCAGCAGTTTCTAGGA AGCCAAAAAATGTAAGTCCAGGCTTCATCAAGATCCAACAGCCTGTAGAAGAGGAGTCGGAGTTGGATT TGGAGCTGGAATCAGAACAAGAGGCTGGGCTGGGTCTGGACCTAGACCTGGATCGGGAGCTGGAGCCTGA AATGGAGCCCAAGGCTGAGACCGAGACCAAGACCCAGACCGAGATGGAGCCCCAGCCTGAGACTGAGCCT GAGATGGAGCCCAACCCCAAATCTAGGCCAAGAGCTCACACTTTTCTCAGCAGCGTTACTGGCCTATGT ATCATCCGCTATGGCTTCCACCCAGTCTCAGACTCAGACTCGGACATGGTCAAGTGGAGAGGAGACGCCA TCCTATGGATTCACTACCCAGAGGGCAACAGCAATGAAGATGTCTAG</p>

In a search of public sequence databases, the NOV51 nucleic acid sequence, located on chromosome 6 has 128 of 198 bases (64%) identical to a gb:GENBANK-ID:AF189262|acc:AF189262.1 mRNA from Rattus norvegicus (Rattus norvegicus GABA-A receptor epsilon-like subunit (Epsilon) mRNA, complete cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV51 polypeptide (SEQ ID NO:120) encoded by SEQ ID NO:119 has 692 amino acid residues and is presented in Table 51B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV51 has a signal peptide and is likely to be localized in the plasma membrane with a certainty of 0.6000. The most likely cleavage site for a NOV51 peptide is between amino acids 69 and 70.

Table 51B. Encoded NOV51 protein sequence (SEQ ID NO:120).

MVAVTDFNSSVHQEYRFQIAYILKTCCKEAMVNVNLTRESSRKGPISWYYLIMGVLG
 LGFIATYLPESAMSAYLAVALHIMLSQLTFIFGIMISFHAGPISFFYDIINYCVLPKA
 NSTSILVFLTUVVALRINKCIRISFNQYPIEFPMELFLIIGFTVIANKISMATETSQTLI
 DMIPYSFLLPVTPDFSLLPKIILQAFSLSLVSSFLLIIFLGKKIASLHNVSVNSNQDLIAI
 GLCNVSSFFRSCVFTGAIARTIIQDKSGGRQQFASLVGAGVMLLLMVKMGHFFFTLPNV
 DMVKVPLKEEIIFSLFNSSDTNLQGGKICRCFCNCDDLEPLPRILYTERFENKLDPEASS
 INLIHCSHFESMNTSQTASEDQVPYTVSSVSQKNQQQYEEVEEVWLPNNSSRNSSPGLP
 DVAESQGRRLIPYSDASLLPSVHTIILDFSMVHYVDSRGLVVLRLQICNAFQANILILI
 AGCHSSIVRAFERNDFFDAGITKTQLFSLVHDAVLFALSRKVIGSSELSIDSETVIRET
 YSETDKNDNSRYKMSSSFLGSQKNVSPGFIKIQPVEESELDLESESEAGLGLDLDL
 DRELEPEMEPKAETETKTQTEMEPPQETEPMEPNPKSRPRAHTFPQQRYPMPYHPSMAS
 TQSQTQTRTWSVERRRHPMDSYSPEGNSNEDV

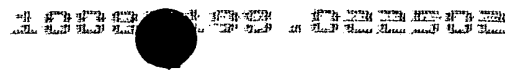
A search of sequence databases reveals that the NOV51 amino acid sequence has 123 of 123 amino acid residues (100%) identical to, and 123 of 123 amino acid residues (100%) similar to, the 123 amino acid residue ptnr:SPTREMBL-ACC:Q9NQP0 protein from Homo sapiens (Human) (BA48209.2 (NOVEL SULPHATE TRANSPORTER FAMILY MEMBER)). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV51 is expressed in at least Adipose, Peripheral Blood, Spinal Chord, Testis, and Colon. Expression information was derived from the tissue sources of the sequences that were included in the derivation of the sequence of CG57785_01. The sequence is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:AF189262|acc:AF189262.1) a closely related Rattus norvegicus GABA-A receptor epsilon-like subunit (Epsilon) mRNA, complete cds homolog in species Rattus norvegicus :Adipose, Peripheral Blood, Spinal Chord, Testis,Colon. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV51 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 51C.

Table 51C. BLAST results for NOV51

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 16418457 ref NP_443193.1 (NM_052961)	solute carrier family 26, member 8 [Homo sapiens]	970	394/396 (99%)	394/396 (99%)	0.0
gi 16588684 gb AAL26868.1 AF314959.1 (AF314959)	anion transporter/exchanger-8 [Homo sapiens]	970	405/464 (87%)	418/464 (89%)	0.0



derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 36 percent of the bases may be so changed.

5 The disclosed NOV51 protein of the invention includes the sulfate transporter-like protein whose sequence is provided in Table 51B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 51B while still encoding a protein that maintains its sulfate transporter-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein,
10 up to about 0 percent of the residues may be so changed.

 The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

 The above defined information for this invention suggests that this sulfate transporter-like protein (NOV51) may function as a member of a “sulfate transporter family”. Therefore,
15 the NOV51 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene
20 delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

 The NOV51 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the sulfate transporter-like
25 protein (NOV51) may be useful in gene therapy, and the sulfate transporter-like protein (NOV51) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from diastrophic dysplasia and certain other skeletal dysplasias, and adenoma, familial chloride diarrhea, CNS disorders, brain disorders including epilepsy,
30 eating disorders, schizophrenia, ADD; cancer; heart disease; inflammation and autoimmune disorders including Crohn’s disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, blood disorders; psoriasis colon cancer, leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, pancreatic disorders including pancreatic insufficiency

CCTTATCATGCATCTCTGCGAAGAACTGCCCCAGAGAATCAGAAGTTCATCTTGCAGGA
GGATGGATCTTTATTTACGAACAGTCCAAGAAATGTGTCCAGGCTGCGAGGAAGGAGTC
GAGTGACAGTTTCGTTCCACTCTTACGAGACTGCACCAACTCGGATCATCAGAAATGGTT
CTTCAAAGAGCGCATGTTATGAAGCCTCGTGTATCAAGGAGCCCCATCGAAGGAGACTGT
GGAGCCAGGACTCTGCCCCAACAAAGACTTAGCTAAGCAGTGACCAGAACCACAAAACTA
GGCTGGATTGCTTTTGCAAGAGGCAATCATTTGCCCTTTGTGAAAGTGTGTGGATTAGGT
AACAGTGATAGCTGTACTATTTGGCACCTTCTAATGTTCAAATACCTATTTCCAGGTACT
CAGATGGTACCCTGTTTTTGAATTAACCTTTAATTTTCTTCAAACGTATTTAACACGCGG
CCTAACTTCTAGACAAGAAAGATCTTCGGGGGTACAAACCCCCGAAGAATTCGGCGGACC
GTCCACCCTGCTACTAGTCAACCGCGGAGCCAACAACGCCAAGCGCTGCATCACACTCTA
GCACGGCGGCCCCACACGAACACATCAAGCAGAGGCAGATACCATAATAGT

In a search of public sequence databases, the NOV52 nucleic acid sequence, located on chromosome has 951 of 1468 bases (64%) identical to a gb:GENBANK-

ID:MMU73819|acc:U73819.1 mRNA from Mus musculus (Mus musculus polypeptide

5 GalNAc transferase-T4 (ppGaNTase-T4) mRNA, complete cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV52 polypeptide (SEQ ID NO:122) encoded by SEQ ID NO:121 has 581 amino acid residues and is presented in Table 52B using the one-letter amino acid code.

10 Signal P, Psort and/or Hydropathy results predict that NOV52 has a signal peptide and is likely to be localized extracellularly with a certainty of 0.8200. The most likely cleavage site for a NOV52 peptide is between amino acids 39 and 40.

Table 52B. Encoded NOV52 protein sequence (SEQ ID NO:122).

MWGRTARRRCPRELRRGREALLVLLALLALAGLGSVLRAQRGAGAGAAEPGPRTPRPGR
REPVMRPPVPANALGARGEAVRLQLQGEELRLQEEVRLHQINIYLSDRISLHRRRLPER
WNPLCKEKKYDYDNLPRTSV I IAFYNEAWSTLLRTVYSVLETSPDILLEEVILVDDYSR
EHLKERLANELSGLPKVR LIRANKREGLVRARLLGASAARGDVLTFDCHCECHEGWLEP
LLQRIHEEESAVVCPVIDVIDWNTFEYLGNSGEPQIGGFDRVLVFTWHTVPERERIRMQS
PVDVIRSPTMAGGLFAVSKKYFEYLGSYDTGMEVWGGENLEFSFRIWQCGGVLETHPCSH
VGHVFPKQAPYSRNKALANSVRAAEVWMEDEFKELYHRNPRARLEPFGDVTERKQLRDKL
QCKDFKWFLETVPYELHVPEDRPGFFGMLQNKGLTDYCFDYNPPDENQIVGHQVILYLCH
GMGQNQFFEYTSQKEIRYNTHQPEGCI AVEAGMDTLIMHLCEETAPENQKFLQEDGSLF
HEQSKKCVQAARKESSDSFVPLLRDCTNSDHQKWFFKERML

A search of sequence databases reveals that the NOV52 amino acid sequence has 330 of 566 amino acid residues (58%) identical to, and 408 of 566 amino acid residues (72%) similar to, the 578 amino acid residue ptnr:SPTREMBL-ACC:O08832 protein from Mus musculus (Mouse) (POLYPEPTIDE GALNAC TRANSFERASE-T4). Public amino acid
15 databases include the GenBank databases, SwissProt, PDB and PIR.

NOV52 is expressed in at least bone marrow, colon, lung, ovary, kidney, respiratory bronchiole, stomach, testis, tonsils, and germ cells. Expression information was derived from
20 the tissue sources of the sequences that were included in the derivation of the sequence of

CG57748-01. The sequence is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:MMU73819|acc:U73819.1) a closely related Mus musculus polypeptide GalNAc transferase-T4 (ppGaNTase-T4) mRNA, complete cds homolog in species Mus musculus :spleen. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV52 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table52C.

10

Table 52C. BLAST results for NOV52					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 7657112 ref NP_056552.1 (NM_015737)	UDP-N-acetyl- alpha-D- galactosamine:poly peptide N- acetylgalactosamin yltransferase 4; ppGaNTase-T4 [Mus musculus]	578	330/570 (57%)	408/570 (70%)	e-180
gi 4503901 ref NP_003765.1 (NM_003774)	polypeptide N- acetylgalactosamin yltransferase 4 UDP-N-acetyl- alpha-D- galactosamine:poly peptide N- acetylgalactosamin yltransferase 4; GalNAc transferase 4; UDP-GalNAc: polypeptide N- acetylgalactosamin yltransferase 4	578	324/570 (56%)	402/570 (69%)	e-179
gi 13375881 ref NP_078918.1 (NM_024642)	protein FLJ21212 [Homo sapiens]	284	284/284 (100%)	284/284 (100%)	e-174
gi 15530299 gb AAH13945.1 AAH13945 (BC013945)	Similar to hypothetical protein FLJ21212 [Homo sapiens]	272	272/272 (100%)	272/272 (100%)	e-166
gi 14530626 emb CA42368.1 (AL110487)	cDNA EST EMBL:AF031835 comes from this gene [Caenorhabditis elegans]	623	248/529 (46%)	322/529 (59%)	e-128

Tables 52D-E list the domain descriptions from DOMAIN analysis results of NOV52. This indicates that the NOV52 sequence has properties similar to those of other proteins known to contain this domain.

15

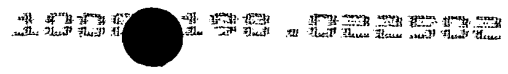


Table 52D. Domain Analysis of NOV52

[gnl|Pfam|pfam00535](#), Glycos_transf_2, Glycosyl transferase. Diverse family, transferring sugar from UDP-glucose, UDP-N-acetyl-galactosamine, GDP-mannose or CDP-abequose, to a range of substrates including cellulose, dolichol phosphate and teichoic acids.

CD-Length = 168 residues, 100.0% aligned

Score = 109 bits (273), Expect = 4e-25

Table 52E. Domain Analysis of NOV52

[gnl|Smart|smart00458](#), RICIN, Ricin-type beta-trefoil; Carbohydrate-binding domain formed from presumed gene triplication

CD-Length = 125 residues, 97.6% aligned

Score = 78.2 bits (191), Expect = 1e-15

NOV52 has homology with uridine diphosphate (UDP)-GalNAc: polypeptide N-acetylgalactosaminyltransferase (GalNAc transferase), a member of the glycosyl transferase family. This enzyme catalyzes the initial step in mucin-type O-glycosylation of specific proteins. Glycosylation of cell surface proteins is critical to normal development, immune response and tissue functions, as evidenced by the phenotypes of a number of mouse knockout models (See Muramatsu; J Biochem (Tokyo) 2000 Feb;127(2):171-6). Glycosylation patterns are known to change during the process of carcinogenesis (Kohsaki et al., J Gastroenterol 2000;35(11):840-8). Alterations of these patterns by introducing a transgene coding for a GalNAc transferase (See Tsurifune et al., Int J Oncol 2000 Jul;17(1):159-65) or by means of antisense oligonucleotides (See Zeng et al., Proc Natl Acad Sci U S A 1995 Sep 12;92(19):8670-4) alter cell morphology, growth and adhesion patterns. Therefore these proteins are important markers and therapeutic targets for oncology applications. In addition, a member of this family has been implicated in autosomal dominant hypophosphatemic rickets (See White et al., Gene 2000 Apr 4;246(1-2):347-56).

Glycosyl transferases comprise a fairly diverse group of proteins that catalyze the addition of sugar from UDP-glucose, UDP-N-acetyl-galactosamine, GDP-mannose or CDP-abequose, to a range of substrates including cellulose, dolichol phosphate and teichoic acids.

The disclosed NOV52 nucleic acid of the invention encoding a N-acetylgalactosaminyltransferase-like protein includes the nucleic acid whose sequence is provided in Table 52A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 52A while still encoding a protein that maintains its N-acetylgalactosaminyltransferase-like

activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or
5 complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.
10 In the mutant or variant nucleic acids, and their complements, up to about 36 percent of the bases may be so changed.

The disclosed NOV52 protein of the invention includes the N-acetylgalactosaminyltransferase-like protein whose sequence is provided in Table 52B. The invention also includes a mutant or variant protein any of whose residues may be changed
15 from the corresponding residue shown in Table 52B while still encoding a protein that maintains its N-acetylgalactosaminyltransferase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 42 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or
20 (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this N-acetylgalactosaminyltransferase-like protein (NOV52) may function as a member of a “N-acetylgalactosaminyltransferase family”. Therefore, the NOV52 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not
25 limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing
30 (but not limited to) those defined here.

The NOV52 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the N-acetylgalactosaminyltransferase-like protein (NOV52) may be useful in gene therapy, and the

N-acetylgalactosaminyltransferase-like protein (NOV52) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, graft versus
 5 host disease, hypercoagulation, autoimmune disease, allergies,transplantation, Hirschsprung's disease , Crohn's Disease, appendicitis, systemic lupus erythematosus, autoimmune disease, asthma, emphysema, scleroderma, allergy, ARDS, endometriosis, fertility, diabetes, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, hypercalceimia, Lesch-Nyhan
 10 syndrome, hypercalceimia, ulcers, fertility, hypogonadism, polycystic ovarian syndrome, cancer, tissue degeneration, bacterial/viral/parasitic infection, or other pathologies or conditions. The NOV52 nucleic acid encoding the N-acetylgalactosaminyltransferase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

15 NOV52 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV52 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV52 proteins have multiple hydrophilic
 20 regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV53

25 A disclosed NOV53 nucleic acid of 2030 nucleotides (also referred to as CG57693-01) encoding a protein kinase-like protein is shown in Table 53A. The start and stop codons are in bold letters.

Table 53A. NOV53 nucleotide sequence (SEQ ID NO:123).

<p>TTTTTTTTTTTTTTGACAATCACCCAGTCAGTATTTATTAGGCGCCTACTGCGGACGGTT GGTCTTTCACGCAGATCAGGCGAGAAGGGATAGTGATGCGCGGGCCGCTGCGGAGACCCGG AGCCCGCCGCGGATCACGGGAATTTTCGCGCCTATTTTTTGGTTGGCTGGGTGTTCTCGCCA GTGATTGGGTCCCGGCAAGGCGTGCCGGTGCGCTCGGCTGCGGCTGCTCCGTCGACCTTT GCAGGACCGGGCGGTGCAGGGCTCACTCGGCTGGCGTCCCGGGGATGGGCCACCAGGAG TCTCCGCTGGCCCGGGCGCCGGCGGGAGGTGCAGCTTATGTAAAGAGGTTATGTAAAGGG CTCAGCTGGCGCGAACACGTGGAAGCCACGGGAGCCTAGGAGCCCAGGCTTCCCCAGCG</p>
--

AGCGCCGCGGCAGCAGAAGGATCCGCTACACGCCGGGCTCGGGCCGCCACCTCCCAGCGCT
GCTCGGTCCCAGGAGGCAGCCCGGGCCCGAGCGGACCATCCCAGGCAGGGGCTCCAGGG
GGGAAACGGGCGCCCGGAAGTGGAGGTGCGCGGGCCAGGTACAATCCAAGGTCCGGCT
CCTCCGCGTCCCAGGGCCGACGAGGGATGAGGCAGGGGGGGCCCGGGCAGCGCCGTTG
CTGCTCCCCCGCGCCCGCAGCCATGGAAACGGGGAAGGACGGCGCCCGCAGAGGTACA
CAAAGCCCGGAGCGGAAAAGGCGAAGCCAGTGCCGCGGGCGCCAGCACGAAGCTGAGG
CCGGCGGCGCGGCCCGGGCCATGGATCCGGTGGCGGGCCGAGGCCCGGGCGAGGCCTTC
CTGGCGCGGCGACGGCCTGAGGGCGGTGGCGGGTCCGCGCGGCCGCTTACAGCCTGTTG
GCGGAGATCGGGCGCGGCAGCTACGGCGTGGTTTATGAGGCAGTGCGCGGGCGCAGCGGG
GCCCCGGGTGGCGGTCAAGAAGATCCGCTGCGACGCCCCGAGAACGTGGAGCTGGCGCTG
GCTGAATTCTGGGCCCTCACCAGCCTCAAGCGGCGCCACCAGAACGTCGTGCAGTTTGAG
GAGTGCCTCTGCAGCGCAATGGGTAGCCAGCGCATGAGTACGGCAACAAGAGCTCG
CAGCTTTACCTGCGCCTGGTGGAGACCTCGCTGAAAGAAAGGATCCTGGGTATGCTGAG
GAGCCCTGCTATCTCTGGTTTGTCTGAGGTTCTGTGAAGGTGAGACCTGAATCAGTAT
GTCCTGTCCCGGAGGCCAGACCCAGCCACCAACAAAAGTTTCATGCTACAGCTGACGAGC
GCCATTGCCTTCTGTCACAAAAACCATATTGTGTCACAGGGACCTGAAGCCAGACAACATC
CTCATCACAGAGCGGTCTGGCACCCCATCCTCAAAGTGGCCGACTTTGGACTAAGCAAG
GTCTGTGCTGGGCTGGCACCCCGAGGCAAGAGGGCAATCAAGACAACAAAAATGTGAAT
GTGAATAAGTACTGGCTGTCTCAGCCTGCGGTTCGGACTTCTACATGGCTCCTGAAGTC
TGGGAGGGACACTACACAGCCAAGGCGGACATCTTTGCCCTGGGCATTATCATCTGGGCA
ATGATAGAAAGAATCACTTTTATTGACTCTGAGACCAAGAAGGAGCTCCTGGGGACCTAC
ATTAAACAGGGGACTGAGATCGTCCCTGTTGGTGAGGCGCTGCTAGAAAACCCAAAGATG
GAGTTCACATCCCCCAAAAACGCAGGACTTCCATGTCTGAGGGGATCAAGCAGCTCTTG
AAAGATATGTTAGCTGCTAACCACAGGACCGGCCTGATGCCCTTGAACCTGAAACCAGA
ATGGACCAGGTACATGTGCTGCTTAAATTCAGGGCTAAGCATTCTTGGGTGATTTTAA
CTAGGTCGATTCTCGGGACCCACAGTCTCACCACGTCTCTCCAGAGGACGGCAGAGGG
TACAGGTGGTGGCCTGGCCGGTTGGCGATCTCCCGACAGCTGGATCCGGC

In a search of public sequence databases, the NOV53 nucleic acid sequence, located on chromosome 20 has 262 of 361 bases (72%) identical to a gb:GENBANK-
ID:AB041802|acc:AB041802.1 mRNA from Mus musculus (Mus musculus brain cDNA,
clone MNCb-1723). Public nucleotide databases include all GenBank databases and the
GeneSeq patent database.

The disclosed NOV53 polypeptide (SEQ ID NO:124) encoded by SEQ ID NO:123 has 533 amino acid residues and is presented in Table 53B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV53 has no signal peptide and is likely to be localized to the cytoplasm with a certainty of 0.8500.

Table 53B. Encoded NOV53 protein sequence (SEQ ID NO:124).

MGHQESPLARAPAGGAAYVKRLCKGLSWREHVESHGSLGAQASPASAAAAGSATRRARA
ATSRAARSRRQPGPGADHPQAGAPGGKRAARKWRCAGQVTIQGPAPPRPRAGRRDEAGGA
RAAPLLLPPPPAAMETGKDGAARRGTQSPERKRRSPVPRAPSTKLRPAAAARAMDPVAAEA
PGEAFLARRRPEGGGGSARPRYSLLAEIGRGSYGVVYEAVAGRSGARVAVKKIRCDAPEN
VELALAEFWALTSKRRHQNVVQFEECVLQRNGLAQRMESHGKSSQLYLRLVETSLKERI
LGYAEPCYLFVMEFCGEGDLNQYVLSRRPDPATNKSFMLQLTSAIAFLKHNHIVHRDL
KPDNILITERSGTPILKVADFGLSKVCAGLAPRGKEGNQDNKNVNVNKYWLSSACGSDFY
MAPEVWEGHYTAKADIFALGIIIWAMIERITFIDSETKKELLGTYIKQGTEIVPVGEALL
ENPKMELHIPQKRRTSMSEGIKQLLKDMLAANPQDRPDAFELETRMDQVTCOA

A search of sequence databases reveals that the NOV53 amino acid sequence has 517 of 517 amino acid residues (100%) identical to, and 517 of 517 amino acid residues (100%)

similar to, the 517 amino acid residue ptnr:TREMBLNEW-ACC:CAC10518 protein from Homo sapiens (Human) (BA55008.2 (NOVEL PROTEIN KINASE)). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV53 is expressed in at least Adrenal Gland/Suprarenal gland, Lymphoid tissue, Oviduct/Uterine Tube/Fallopian tube, Peripheral Blood, Placenta, Retina, Thymus. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV53 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 53C .

Table 53C. BLAST results for NOV53					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 12830335 emb CAC10518.2 (AL359916)	ba55008.2 (novel protein kinase) [Homo sapiens]	517	517/517 (100%)	517/517 (100%)	0.0
gi 18592261 ref XP_086681.1 (XM_086681)	serine/threonine kinase 35 [Homo sapiens]	401	400/401 (99%)	400/401 (99%)	0.0
gi 16878290 gb AAH17340.1 AAH17340 (BC017340)	(protein for IMAGE:4869353) [Homo sapiens]	398	395/398 (99%)	395/398 (99%)	0.0
gi 18549074 ref XP_086530.1 (XM_086530)	similar to Cell division control protein 2 homolog (P34 protein kinase) [Homo sapiens]	161	101/156 (64%)	129/156 (81%)	e-56
gi 15224378 ref NP_181320.1 (NM_129340)	putative protein kinase [Arabidopsis thaliana]	257	92/314 (29%)	143/314 (45%)	e-23

Tables 53D-F list the domain descriptions from DOMAIN analysis results against NOV53. This indicates that the NOV53 sequence has properties similar to those of other proteins known to contain this domain.

Table 53D. Domain Analysis of NOV53
gnl Smart smart00220 , S_TKc, Serine/Threonine protein kinases, catalytic domain; Phosphotransferases. Serine or threonine-specific kinase subfamily.
CD-Length = 256 residues, 97.7% aligned
Score = 193 bits (491), Expect = 2e-50

[illegible]

CD-Length = 256 residues, 97.7% aligned

301

Table 53F. Domain Analysis of NOV53

gnl|Smart|smart00219, TyrKc, Tyrosine kinase, catalytic domain;
Phosphotransferases. Tyrosine-specific kinase subfamily

CD-Length = 258 residues, 99.6% aligned

Score = 121 bits (303), Expect = 1e-28

Protein phosphorylation is a fundamental process for the regulation of cellular functions. The coordinated action of both protein kinases and phosphatases controls the levels of phosphorylation and, hence, the activity of specific target proteins. One of the predominant roles of protein phosphorylation is in signal transduction, where extracellular signals are amplified and propagated by a cascade of protein phosphorylation and dephosphorylation events. Eukaryotic protein kinases are enzymes that belong to a very extensive family of proteins which share a conserved catalytic core common with both serine/threonine and tyrosine protein kinases. There are a number of conserved regions in the catalytic domain of protein kinases. In the N-terminal extremity of the catalytic domain there is a glycine-rich stretch of residues in the vicinity of a lysine residue, which has been shown to be involved in ATP binding. In the central part of the catalytic domain there is a conserved aspartic acid residue which is important for the catalytic activity of the enzyme. Protein kinases are excellent small molecule drug targets for therapeutic intervention.

The disclosed NOV53 nucleic acid of the invention encoding a protein kinase-like protein includes the nucleic acid whose sequence is provided in Table 53A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 53A while still encoding a protein that maintains its protein kinase-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 28 percent of the bases may be so changed.

The disclosed NOV53 protein of the invention includes the protein kinase-like protein whose sequence is provided in Table 53B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 53B while still encoding a protein that maintains its protein kinase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 0 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this protein kinase-like protein (NOV53) may function as a member of a "protein kinase family". Therefore, the NOV53 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV53 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the protein kinase-like protein (NOV53) may be useful in gene therapy, and the protein kinase-like protein (NOV53) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from adrenoleukodystrophy, congenital adrenal hyperplasia, anemia, ataxia-telangiectasia, autoimmune disease, fertility, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host disease, allergies, immunodeficiencies, transplantation, graft versus host disease (GVHD), lymphoedema, Von Hippel-Lindau (VHL) syndrome, diabetes, tuberous sclerosis, or other pathologies or conditions. The NOV53 nucleic acid encoding the protein kinase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV53 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV53 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods

known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV53 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV54

A disclosed NOV54 nucleic acid of 3331 nucleotides (also referred to as CG57707-01) encoding a Leucine-rich glioma-inactivated protein precursor -like protein is shown in Table 54A. The start and stop codons are in bold letters.

Table 54A. NOV54 nucleotide sequence (SEQ ID NO:125).

ATGCGCTGCGGAGAGGCGGCTGCGGAGCGCTCGGGCTGCTGCTGCTGCTGCTGGGCGCC
GCGTGCTGATACCGCGGAGCGCGCAGGTGAGGCGGCTGGCGCGCTGCCCCGCCACTTGC
AGCTGTACCAAGGAGTCTATCATCTGCGTGGGCTCTTCTGGGTGCCAGGATCGTGCCG
GGCGACATCAGCTCCCTGAGCCTGGTAAATGGGACGTTTTTCAGAAATCAAGGACCGAATG
TTTCCCATCTGCCTTCTCTGCAGCTGCTATTGCTGAATTCTAACTCATTCACGATCATC
CGGGATGATGCTTTTGCTGGACTTTTTTCATCTTGAATACCTGTTCAATTGAAGGGAACAAA
ATAGAAACCATTTCAAGAAATGCCTTTTCGTGGCCTCCGTGACCTGACTCACCTTTCTTTG
GCCAATAACCATATAAAGCACTACCAAGGGATGTCTTCAGTGATTAGACTCTCTGATT
GAACAGATTTTGAGGGGTAATAAATTTGAATGTGACTGCAAAGCCAAGTGGCTATACCTG
TGTTTGAAGATGACAAATTCACCGTTTTCTGATGTGCTGTGTATTGGTCCACCAGAGTAT
CAGGAAAAGAAGCTAAATGACGTGACCAGCTTTGACTATGAATGCACAACTACAGATTTT
GTTGTTTCATCAGACTTTACCCTACCAGTCGGTTTTCAGTGGATACGTTCAACTCCAAGAAC
GATGTGTACGTGGCCATCGCGCAGCCAGCATGGAGAATGCGATGGTGCTGGAGTGGGAC
CACATTGAAATGAATTTCCGGAGCTATGACAACATTACAGGTGCTAGTCCATCGTGGGCTGT
AAGGCCATTCTCATCGATGATCAGGTCTTTGTGGTGGTAGCCAGCTCTTCGGTGGCTCT
CATTTTACAAATACGACGAGAGTTGGACCAATTTGTCAAATTTCAAGACATAGAGGTC
TCTCGCATTTTCCAAGCCCAATGACATCGAGCTGTTTCAGATCGACGACGAGACGTTCTTT
GTCATCGCAGACAGCTCAAAGGCTGGTCTGTCCACAGTTTATAAATGGAACAGCAAAGGA
TTCTATTCTTACCAGTCACTGCACGAGTGGTTTCAAGGACACGGATGCGGAGTTTGTGAT
ATCGATGGAATTCGATCTCATCTGTCCAGCCGCTCCAGGTCCCATCATCTCCAG
TGGAATAAAGCTCTAAGAAGTTTGTCCCCCATGGTGACATCCCCAACATGGAGGACGTA
CTGGCTGTGAAGAGCTTCCGAATGCAAATACCTCTACCTTTCCCTTACCCGCTTCATC
GGGACTCCCGGGTCATGAGGTGGAACAGTAAGCAGTTTGTGGAGATCCAAGCTCTTCCA
TCCCGGGGGGCCATGACCCTGCAGCCCTTTTCTTTTAAAGATAATCACTACCTGGCCCTG
GGGAGTGACTATACATTCTCTCAGATATACCAAGTGGGATAAAGAGAAGCAGCTATTCAAA
AAGTTTAAGGAGATTTATGTGCAGGCGCCTCGTTCATTACAGCTGTCTCCACCGACAGG
AGAGATTTCTTTTTCATCCAGTTTCAAAGGGAAAACAAAGATTTTGAACATATAATT
GTTGACTTAAGTTTGTGAAGGTGTGGTGGGTGAACTAAGAGAAATGTAGCATTAGCTCT
CACAAAAGAGGACCAAGAAAAATCAACAAACAAATCAAAGCCAGGCTCAGAGCTCTGAAA
TTAAAAGCACTGAAATAGTTAGATGTTTTCAAACTTTGTAGAACTCACATTTTAATCAGG
GATTGCATTTATTGGCTAACTGCATGACATGCCATTCTACCATTTTAAAAAATCTTA
AAGCCTGTAATTTCTGAGAAAAGAGTACAGCATTTACTCTTATCATCTAGAAATGTAATA
TGCTTCCCCCGCTTTTGTGATGAGGAAGAAGACAATTGGATAAGATGGGACAGCACTTA
TAATGAAATAAAAAAACTTTGAGCCCTCTCATTCCACTTTAGCAATCTTTTGGTAA
GAACTCTTAAAGCCAAAAGTCTGCTGAAAAGATTGCTGATTATTAGTTTAAAAATCTTG
TAACACTCAGCAGTGCTATTTTGTGATCATCCAGTTTCTGAAAGTAATGCCAGTCTTC
CTGAATCTCTTAATAGCAGAACCTTGGTGATTTGTTGGCTCATATGAATGCTTGTC
TGGATATGTTAACAATTTAGTGTGTTGACATTGCTTCTCTGCCACAAAGACAATACTCTG

GTGACACATGTCTAGGCCAGCACAGGCTGTAGGCCAGGAGTGAAGGAGTTTTT
 CCTCTTTCTTACGGTTCAAAGGTGACCCTGGTGGTGGCCAGAGCAGTAATGCTTGTG
 ATGCTCTTCATGGCTCATCTGCTTCTCAGAACCCACCCGTTGAGTTTGTGGGTAACCAGC
 AGGCAGGCCAAAGACTGGTGCTTTTCATTTTCATCCTTTAGAGGGATGAAACAGTTATTTT
 CGTCTGATGAGCATTCCGGTAGAATTTTTGAAGTGAGATTTTATGAAGTCAAAGGGGACTT
 TACACAGATCTCGACCTGCTTTGAAACCTAGAGGTGGCCCTTTGATTTGTGCGTGTCCTT
 GCCCTCTGGACAACTTAATATTTCAAGTAATCGAATACCAACTTCCCTGCCAGCCCACCT
 GCCTTCCGCCCCGCTTGTGTAACAGTCTGTTTTGTTGAGTTGCTGCTATTGCACTGCCA
 GTGCAGCCACACCAAATCACAAACCAAGATACTCAGATAGGAAGACTCCTTCTCTCCC
 AGTACTTTACCAAAGGAACCCCCGCCAGGACCCACATGGGGCCACGTGTTGGCAGTGGAA
 TCAGCTGTGCAGGCTGGGGATCTCAGGCTGATCAGTAGGGGCCAGCTTTGGAGCCAGCC
 AAGCTGAATCCCACTCCAGGTCTGTGCTCAAGAGACCAGATGGTGTATTTCCAAATGG
 GCCTCTCTGGTATGGGCAATAGGCAAGCTCCTGGGGTCTGGTTATGTGGAAGATTCTTAG
 TGGATGTTCCGCCTGGTTAGCTGGTTCTCTTCAGAGAATATAAAGTGAATGCCTTTAGGG
 GTAGCTCTGAAAGAGAAACCCAAACAATTCATTCTAGCCATGAAAGTAGCAGATCATA
 TTGTAATGTTATTGTTAAATGACTATTTGCCATGTCATGAGTAGGTAGATGTTTT
 GCCACAAATATGAATGTGTTTGTGTTCTGACTTTAAGCAATGAAGATTGAGACAATA
 AATAGCACTCAGAGAATGAAGCATTGATGTT

In a search of public sequence databases, the NOV54 nucleic acid sequence, located on chromosome 4p16 has 682 of 1052 bases (64%) identical to a gb:GENBANK-ID:AF055636|acc:AF055636.1 mRNA from Homo sapiens (Homo sapiens leucine-rich glioma-inactivated protein precursor (LGI1) mRNA, complete cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV54 polypeptide (SEQ ID NO:126) encoded by SEQ ID NO:125 has 545 amino acid residues and is presented in Table 54B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV54 has a signal peptide and is likely to be localized localized extracellularly with a certainty of 0.8200. The most likely cleavage site for a NOV54 peptide is between amino acids 22 and 23.

Table 54B. Encoded NOV54 protein sequence (SEQ ID NO:126).

MALRRGGCGALGLLLLLLGAACLI PRSAQVRR LARCPATCSCTKESIICVGSSWVPRIVP
 GDISLSLVNGTFSEIKDRMFSLPSLQLLLNSNSFTIIRDDAFAGLFHLEYLFIEGNK
 IETISRNAFRGLRDLTHLSLANNHIKALPRDVFSDLSLIEQILRGNKFECDCKAKWLYL
 WLKMTNSTVSDVLCIGPPEYQEKKLNVDVTSFDYECTTTDFVHQTLPYQSVSDTFNSKN
 DVYVAIAQPSMENC MVLEWDHIEMNFRSYDNITGQSIVGCKAILIDDQVFVVVAQLFGGS
 HIYKYDESWTKFVKFQDIEVSRISKPNIDIELFQIDDETFVIADSSKAGLSTVYKWNKSG
 FYSYQSLHEWFRD TDAEFVDIDGKSHLILSSRSQVPIILQWNKSSKKFVPHGDIPNMDV
 LAVKSFRMQNTLYLSLTRFIGDSRVMRWNSKQFVEIQALPSRGAMTLQPF SFDKNHYLAL
 GSDYTFSQIYQWDKEKQLFKKFKEIYVQAPRSFTAVSTDRRDFFFASSFKGKTKIFEHII
 VDL SL

A search of sequence databases reveals that the NOV54 amino acid sequence has 301 of 538 amino acid residues (55%) identical to, and 386 of 538 amino acid residues (71%) similar to, the 557 amino acid residue ptnr:SPTREMBL-ACC:O95970 protein from Homo sapiens (Human) (LEUCINE-RICH GLIOMA-INACTIVATED PROTEIN PRECURSOR). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

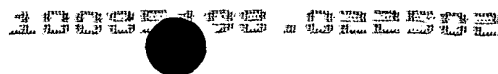
- 5 The disclosed NOV54 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 54C.

Table 54C. BLAST results for NOV54					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 15620891 dbj BAB67809.1 (AB067503)	KIAA1916 protein [Homo sapiens]	542	539/542 (99%)	539/542 (99%)	0.0
gi 9938002 ref NP_064674.1 (NM_020278)	leucine-rich, glioma inactivated 1 [Mus musculus]	557	296/516 (57%)	378/516 (72%)	e-178
gi 4826816 ref NP_005088.1 (NM_005097)	leucine-rich, glioma inactivated 1 precursor [Homo sapiens]	557	296/516 (57%)	379/516 (73%)	e-178
gi 15722102 emb CAC78757.1 (AL358154)	bA512J3.1 (leucine-rich, glioma inactivated 1) [Homo sapiens]	461	269/460 (58%)	342/460 (73%)	e-160
gi 18591028 ref XP_092048.1 (XM_092048)	protein XP_092048 [Homo sapiens]	466	179/437 (40%)	267/437 (60%)	3e-93

- 10 Table 54D lists the domain descriptions from DOMAIN analysis results against NOV54. This indicates that the NOV54 sequence has properties similar to those of other proteins known to contain this domain.

Table 54D. Domain Analysis of NOV54
<p>gnl Pfam pfam01463, LRRCT, Leucine rich repeat C-terminal domain. Leucine Rich Repeats pfam00560 are short sequence motifs present in a number of proteins with diverse functions and cellular locations. Leucine Rich Repeats are often flanked by cysteine rich domains. This domain is often found at the C-terminus of tandem leucine rich repeats.</p> <p>CD-Length = 51 residues, 98.0% aligned, Score = 43.1 bits (100), Expect = 4e-05</p>

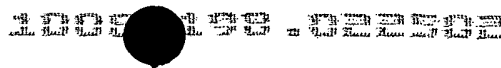
- 15 Loss of heterozygosity for 10q23-26 is seen in over 80% of glioblastoma multiforme tumors. Positional cloning was used to isolate the LGI1 (Leucine-rich gene-Glioma



Inactivated) gene, which is rearranged as a result of the t(10;19)(q24;q13) balanced translocation in the T98G glioblastoma cell line lacking any normal chromosome 10 (See Chernova et al., *Oncogene* 17: 2873-2881). Rearrangement of the LGI1 gene was also detected in the A172 glioblastoma cell line and several glioblastoma tumors. These rearrangements lead to a complete absence of LGI1 expression in glioblastoma cells. The LGI1 gene encodes a protein with a calculated molecular mass of 60 kD and contains 3.5 leucine-rich repeats (LRR) with conserved flanking sequences. In the LRR domain, LGI1 has the highest homology with a number of transmembrane and extracellular proteins which function as receptors and adhesion proteins. LGI1 is predominantly expressed in neural tissues, especially in brain; its expression is reduced in low grade brain tumors and it is significantly reduced or absent in malignant gliomas. Its localization to the 10q24 region, and rearrangements or inactivation in malignant brain tumors, suggest that LGI1 is a candidate tumor suppressor gene involved in progression of glial tumors.

The human leucine-rich glioma-inactivated protein precursor-like protein described in this invention is predicted to share the attributes of other family members and is thus implicated in regulation of cell growth and survival as well as cellular metabolism. Like the LGI1 gene, the leucine-rich glioma-inactivated protein precursor-like gene described in this patent is expressed in neural tissues; however, it also appears to be frequently expressed in parathyroid tumors. Therefore, this protein is an attractive target for drug intervention in the treatment of cancer, central nervous system disorders, and metabolic diseases, among others. The leucine-rich glioma-inactivated protein precursor-like gene maps to human chromosome 4p16 and is predicted to encode a secreted protein.

The disclosed NOV54 nucleic acid of the invention encoding a Leucine-rich glioma-inactivated protein precursor-like protein includes the nucleic acid whose sequence is provided in Table 54A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 54A while still encoding a protein that maintains its Leucine-rich glioma-inactivated protein precursor-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least



in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 36 percent of the bases may be so changed.

5 The disclosed NOV54 protein of the invention includes the Leucine-rich glioma-inactivated protein precursor-like protein whose sequence is provided in Table 54B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 54B while still encoding a protein that maintains its Leucine-rich glioma-inactivated protein precursor-like activities and
10 physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 45 percent of the residues may be so changed.

 The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

 The above defined information for this invention suggests that this Leucine-rich
15 glioma-inactivated protein precursor-like protein (NOV54) may function as a member of a “Leucine-rich glioma-inactivated protein precursor family”. Therefore, the NOV54 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein
20 therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

 The NOV54 nucleic acids and proteins of the invention are useful in potential
25 therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the Leucine-rich glioma-inactivated protein precursor-like protein (NOV54) may be useful in gene therapy, and the Leucine-rich glioma-inactivated protein precursor-like protein (NOV54) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions
30 of the present invention will have efficacy for treatment of patients suffering from cancer, trauma, bacterial and viral infections, *in vitro* and *in vivo* regeneration, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety,

pain, neurodegeneration, anemia, bleeding disorders, scleroderma, transplantation, hyperparathyroidism, hypoparathyroidism, diabetes, autoimmune disease, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, Lesch-Nyhan syndrome, Hirschsprung's disease, Crohn's Disease, appendicitis, endometriosis, fertility, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, obesity, transplantation, and pancreatitis, or other pathologies or conditions. The NOV54 nucleic acid encoding the Leucine-rich glioma-inactivated protein precursor-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV54 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV54 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV54 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV55

A disclosed NOV55 nucleic acid of 2886 nucleotides (also referred to as CG57306-01) encoding an anion exchanger-like protein is shown in Table 55A. The start and stop codons are in bold letters.

Table 55A. NOV55 nucleotide sequence (SEQ ID NO:127).

ATTCTGTGCAAGCCT**CAT**GGAAATGAAGCTGCCAGGCCAGGAAGGGTTGAAGCCTCCAG
TGCTCCTAGAAATATTCCTTCAGGGGAGCTGGACAGCAACCCTGACCCTGGCACCGGCC
CAGCCCTGATGGCCCTCAGACACAGAGAGCAAGGAAGTGGGAGTACCCAAGACCCTCT
GCTCTTCATTGAGCTGAATGAGCTGCTGGGCTGGCCCAAGCGCTGGAGTGGAGAGAGAC
AGGCCGATGGGTACTGTTTGGAGAGAAGTTGGAGGTGGCTGCAGGCCGGTGGAGTGCCCC
CCACGTGCCCAACCCTGGCACTGCCAGCCTCCAGAAGCTCCGCAGCCTGCTGGCCGAGGG
CCTTGTAAGTCTGGAGTGGCCAGCTCAGAGCCTCCTGGAGCTCGTGGAGCAGGTGACCAG
GGTGGAGTCGCTGAGCCAGAGCTGAGAGGGCAGTTGCAGGCCTTGTGCTGCAGAGACC
CCAGCATTACAACCAGACCAGGCACAGGCCCTGCTGGGGTGGAGCCCCCTCCAGAAA
GGCTTCTGACAATGAGGAAGCCCCCTGAGGGACCAGTGTGAGAACCCCTGAGACAGAA

GCTACCTCCAGGAGCTGAGGCAGGGACTGTGCTGGCAGGGGAGCTGGGCTTCCTGGCACA
 GCCACTGGGAGCCTTTGTTGACTGCGGAACCTGTGGTACTGGGGTCCCTTACTGAGGT
 GTCCCTCCCAAGCAGGTTTCTGCTTCTCCTGGGCCCTGTATGCTGGGAAAGGGCTA
 CCATGAGATGGGACGGGACAGCTGTCTCTCAGTGACCCGATTCCCAGCAATTCCA
 GTGGTCAGTTCGTCGGGCCAGCAACCTTCATGACCTTCTGGCAGCCCTGGATGCATTCCT
 AGAGGAGGTGACAGTGCTTCCCCAGGTGCGGTGGGACCCAAAGCCCGATTCCCCGCC
 CAAATGTCTGCCATCTCAGCACAAAGGACCTCGGCTGAGGACAGGCACCGCCATGGGCC
 ACACGCACACAGCCCGGAGTTGCAGCGGACCGGCAGGCTGTTTGGGGGCCTTATCCAGGA
 CGTGCGCAGGAAGGTCCCGTGGTACCCAGCGATTTCCTGGACGCCCTGCATCTCCAGTG
 CTTCTCGGCCGTACTCTACATTTACCTGGCCACTGTACTAATGCCATCACTTTTGGGGG
 TCTGCTGGGAGATGCCACTGATGGTGCCAGGGAGTGCTGGAAAGTTTCTGGGCACAGC
 AGTGGCTGGAGTGCCTTCTGCCTGATGGCAGGCCAGCCCTCACCATTCTGAGCAGCAC
 GGGGCCAGTGCTGGTCTTTGAGCGCCTGCTCTTCTTTTCTCAGCAGAGATTACAGCCTGGA
 CTACCTGCCCTTCCGCCATGAGGTGGGCATCTGGGTGGCTACCTTTTGCCTGGTGTGGT
 GGCCACAGAGGCCAGTGCTGGTGGCTACTTACCCGCTTCACTGAGGAAGGTTTCTG
 TGCCCTCATCAGCCTCATCTTCATCTACGATGCTGTGGGCAAAATGCTGAACCTGACCCA
 TACCTATCCTATCCAGAAGCCTGGGTCTCTGCCACGGGTGCCTCTGCCAATACCCAGG
 CCCAGGAGGTAATGAGTCTCAATGGATAAGGACAAGGCCAAAAGACAGAGACGACATTGT
 AAGCATGGACTTAGGCCGTGATCAATGCATCCTTGCTGCCGCCACCTGAGTGCACCCGGCA
 GGGAGGCCACCTCGTGGCCCTGGCTGTCTACAGTCCCAGACATTGCCTTCTTCTCCCT
 TCTCCTCTTCTTACTTCTTCTTCTTCTTCTGCTATGGCCCTCAAGTGTGTAAAGACCAGCCG
 CTTCTTCCCCCTCTGTGGTGCGCAAAGGGCTCAGCGACTTCTCCTCAGTCTTGGCCATCCT
 GCTCGGCTGTGGCCTTGATGCTTTCCTGGGCCCTAGCCACACCAAAGCTCATGGTACCCAG
 AGAGTTCAAGCCCACTCCCTGGGCGTGGCTGGCTGGTGTACCTTTTGGAGCCAACCC
 CTGGTGGTGGAGTGTGGCAGCTGCCCTGCCTGCCCTGCTGCTGTCTATCCTCATCTTCAT
 GGACCAACAGATCACAGCAGTCATCCTCAACCGCATGGAATACAGACTGCAGAAGGGAGC
 TGGCTTCCACCTGGACCTCTTCTGTGTGGCTGTGCTGATGCTACTCACATCAGCGCTTGG
 ACTGCCTTGGTATGTCTCAGCCACTGTCTCCTGGCTCACATGGACAGTCTTCGGAG
 AGAGAGCAGAGCCTGTGCCCCCGGGGAGCGCCCCAACTTCTGGGTATCAGGGAACAGAG
 GCTGACAGGCCTGGTGGTGTTCATCCTTACAGGAGCCTCCATCTTCTGGCACCTGTGCT
 CAAGTTCAATCCAATGCCTGTGCTCTATGGCATCTTCTGTATATGGGGGTGGCAGCGCT
 CAGCATTCAGTTCACCTAATAGGGTGAAGCTGTTGTTGATGCCAGCAAACACCCAGCC
 AGACCTGCTACTCTTGCGGCATGTGCCTCTGACCAGGGTCCACCTCTTACAGCCATCCA
 GCTTGCTGTCTGGGGCTGCTTTGGATAATCAAGTCTACCCCTGCAGCCATCATCTTCCC
 CCTCATGTTGCTGGGCCCTGTGGGGGTCCGAAAGGCCCTGGAGAGGGTCTTCTCACCACA
 GGAACCTCTCTGGCTGGATGAGCTGATGCCAGAGGAGGAGAGAAGCATCCCTGAGAAGGG
 GCTGGAGCCAGAACCTCATTAGTGGAAGTGACAGTGAAGATTAGAGCTGATGTATCA
 GCCAAAGGCTCCAGAAATCAACATTTCTGTGAATTAGCTGGAGTAGGAGTCTGGGAGTGG
 AGACCC

In a search of public sequence databases, the NOV nucleic acid sequence, located on
 chromosome 17 has 2250 of 2788 bases (80%) identical to a gb:GENBANK-
 ID:AB038264|acc:AB038264.1 mRNA from *Oryctolagus cuniculus* (*Oryctolagus cuniculus*
 5 AE4b mRNA for anion exchanger 4b, complete cds). Public nucleotide databases include all
 GenBank databases and the GeneSeq patent database.

The disclosed NOV55 polypeptide (SEQ ID NO:128) encoded by SEQ ID NO:127
 has 946 amino acid residues and is presented in Table 55B using the one-letter amino acid
 code. Signal P, Psort and/or Hydropathy results predict that NOV55 has no signal peptide and
 10 is likely to be localized extracellularly with a certainty of 0.8000.

Table 55B. Encoded NOV55 protein sequence (SEQ ID NO:128).
MEMKLPQGFEASSAPRNIPSGELDSNPDPGTGSPDGPSTESKELGVPKDPLLFQL

NELLGWPQALEWRETGRWVLFEEKLEVAAGRWSAPHVPTLALPSLQKLRSLLAEGVLVLLD
CPAQSLLELVEQVTRVESLSPELRGQLQALLLQRPQHYNQTTGTRPCWGESPSRKASDNE
EAPLRDQCQNPLRQKLPPGAEAGTVLAGELGFLAQPLGAFVRLRNPVVLGSLTEVSLPSR
FFCLLLGPCMLGKGYHEMGRAAAVLLSDPHSQQFQWSVRRASNLHDLAALDAFLEEVTV
LPPGRWDPTARI PPPKCLPSQHKRTSAEDRHRHGPHAHSPQLQRTGRLFGGLIQDVRKRV
PWYPSDFLDALHLQCFSVAVLYIYLATVTNAITFGGLLGDATDGAQGVLESFLGTAVAGAA
FCLMAGQPLTILSSTGPPVLVFERLLFSFSRDYSLDYLFPRLWVGIIWVATFCLVLVATEAS
VLVRYFTRFTEEGFCALISLIFIYDAVGKMLNLTHTYPIQKPGSSAYGCLCQYPPGPGNE
SQWIRTRPKDRDDIVSMDLGLINASLLPPECTRQGGHPRGPGCHTVPDIAFFSLLLFLT
SFFFAMALKCCKTSRFFPSVVRKGLSDFSSVLAAILLGCGLDAFLGLATPKLMVPREFKPT
LPGRGWLVSPPGANPWWWSVAAALPALLLSILIFMDQQITAVILNRMEYRLQKGAGFHLD
LFCVAVLMLLTSALGLPWVVSATVISLAHMDSLRRESRACAPGERPNFLGIREQRLTGLV
VFILTGASIFLAPVLKFI PMPVLYGIFLYMGVAALSSIQFTNRVKLLMPAKHQPDLLLL
RHPVPLTRVHLFTAIQLACLGLLWIIKSTPAAIIFPLMLLGLVGVRKALERVFSPPQELLWL
DELMPEEERSIPEKGLEPEHSFSGSDSESELMYQPKAPEINISVN

A search of sequence databases reveals that the NOV55 amino acid sequence has 827 of 944 amino acid residues (87%) identical to, and 873 of 944 amino acid residues (92%) similar to, the 939 amino acid residue ptnr:TREMBLNEW-ACC:BAB18936 protein from *Oryctolagus cuniculus* (Rabbit) (ANION EXCHANGER 4B). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV55 is expressed in at least kidney, testis. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV55 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 55C.

Table 55C. BLAST results for NOV55					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 18560973 ref XP_038736.2 (XM_038736)	solute carrier family 4, sodium bicarbonate cotransporter, member 9 [Homo sapiens]	959	938/962 (97%)	940/962 (97%)	0.0
gi 14582760 gb AAK69625.1 AF332961.1 (AF332961)	anion exchanger AE4 [Homo sapiens]	959	937/962 (97%)	939/962 (97%)	0.0
gi 7363254 dbj BAA93010.1 (AB032762)	sodium bicarbonate cotransporter 5 [Homo sapiens]	957	936/960 (97%)	938/960 (97%)	0.0
gi 13517508 gb AAK28832.1 AF313465.1 (AF313465)	sodium bicarbonate cotransporter [Homo sapiens]	990	938/986 (95%)	940/986 (95%)	0.0
gi 13249295 gb AAK16733.1 AF336237.1 (AF336237)	anion exchanger AE4 [Homo sapiens]	945	922/962 (95%)	926/962 (95%)	0.0

Table 55D lists the domain descriptions from DOMAIN analysis results against NOV55. This indicates that the NOV55 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 55D. Domain Analysis of NOV55	
gnl Pfam pfam00955, HCO3_cotransp, HCO3- transporter family. This family contains Band 3 anion exchange proteins that exchange CL-/HCO3-. This family also includes cotransporters of Na+/HCO3-.	
CD-Length = 781 residues, 100.0% aligned	
Score = 731 bits (1887), Expect = 0.0	

The rabbit anion exchanger 4B protein is a member of the bicarbonate ion transporter superfamily and is present on the apical membrane of beta-intercalated cells in the collecting ducts of the rabbit kidney (See Tsuganezawa et al., J Biol Chem 2000 Dec 1). The rabbit protein has sodium-independent anion exchanger activity when expressed in cultured COS-7 cells and Xenopus oocytes.

The acid-secreting alpha intercalated cells and bicarbonate-secreting beta intercalated cells are sites for modulation of urinary acid secretion, which in turn governs acid-base homeostasis. Mutations in the red cell anion exchanger gene, for instance, are correlated with familial distal renal tubular acidosis (See Bruce et al., Biochem Cell Biol 1998; 76(5): 723-728).

The disclosed NOV55 nucleic acid of the invention encoding a anion exchanger-like protein includes the nucleic acid whose sequence is provided in Table 55A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 55A while still encoding a protein that maintains its anion exchanger-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical

stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 20 percent of the bases may be so changed.

5 The disclosed NOV55 protein of the invention includes the anion exchanger-like protein whose sequence is provided in Table 55B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 55B while still encoding a protein that maintains its anion exchanger-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 13 percent of the residues may be so changed.

10 The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this anion exchanger-like protein (NOV55) may function as a member of a “anion exchanger family”. Therefore, the NOV55 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

20 The NOV55 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the anion exchanger-like protein (NOV55) may be useful in gene therapy, and the anion exchanger-like protein (NOV55) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from acidosis, alkalosis, diabetes, autoimmune disease, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, hypercalcaemia, Lesch-Nyhan syndrome, cancer, tissue degeneration, bacterial/viral/parasitic infection, or other pathologies or conditions. The NOV55 nucleic acid encoding the anion exchanger-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV55 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV55 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV55 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV56

A disclosed NOV56 nucleic acid of 1083 nucleotides (also referred to as CG57348-01) encoding a PR_SET_domain protein-like protein is shown in Table 56A. The start and stop codons are in bold letters.

Table 56A. NOV56 nucleotide sequence (SEQ ID NO:129).

```
GGCGGCGAGGGTGC GGCGCGCGCTGCCATGGGCGCTTGGCGGCGCCCGCGGCGCTGGTGGCGGCGGCAG
CAGCAGCAGCAGCGGCAGCGGCAGCGGTGGTGGCCGGGCGCGGCGGCGGCGGCGAGGGTGC GGCGCGCG
CTGCCATGGGCGCTGGCGGGCTGCAGGCAAGAAGATGTCCAAGCCCCGCGCGCTGGAGGCGGCGGCGGCG
GCGGCAGCGACGGCCCCGGGCGCTGGAGATGGTGGAGCGGAGGGGCGCGGGGAGGCCCGCACCGATGGGG
AGAGCGTATTTACGGGGCAGTCAAAGATCTATTCTACATGAGCCCCGAACAAATGCTCTGGAATGCGTTT
CCCCCTTCAAGAAGAGAACTCGGTTACACATCACGAAGTCAAATGCCAGGGGAAACCATTAGCCGGAATC
TACAGGAAACGAGAAGAGAAAAGAAATACTGGGAACGCAGTACAGAGCGCCATGAAGTCCAAGAAACAGA
AGATCAAAGACGCCAGGAGAGGTCCCTGCAAGGAAAAACACAACAGAATCACAACTTACGGATTTCTA
CCCTGTCCGAAGGAGATCCAGGAAGAGCAAAGCCGAGCTGCAGTCTGAAGAAAGGAAAAGAATAGATGAA
TTGATTGAAAGTGGGAAGGAAGAAGGAATGAAGATTGACCTCATCGATGGCAAAGGCAGGGGTGTGATTG
CCACCAAGCATTTCTCCGGGGTGCTTTGTGGTGAATACCACGGGGACCTCATCGAGATCACCGACGC
CAAGAAACGGGAGGCTCTGTATGCACAGGACCTTCCACGGGCTGCTACATGTACTATTTTCAGTATCTG
AGCAAAACCTACTGGATGCAACTAGAGAGACAAATCGCCAGGAAGACCGATCAATCACAGCAAAAT
GTGGGAACTGCCAAACCAAATGCACGACATCGACGGCGTACCTCACCTCATCCTCATCGCCTCCCAAGA
CATCGCGGCTGGGGAGGAGCTCCTGTATGACTATGGGACCGCAGCAAGGCTTCCATTGAAGCCACCCG
TGGCTGAAGCATTAACCGGTGGGCCCCGCGCCC
```

In a search of public sequence databases, the NOV56 nucleic acid sequence, located on chromosome 12 has 1036 of 1086 bases (95%) identical to a gb:GENBANK-ID:AF287261|acc:AF287261.1 mRNA from Homo sapiens (Homo sapiens PR/SET domain containing protein 07 (SET07) mRNA, complete cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV56 polypeptide (SEQ ID NO:130) encoded by SEQ ID NO:129 has 345 amino acid residues and is presented in Table 56B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV56 has a signal peptide and is

likely to be localized in the endoplasmic reticulum with a certainty of 0.600. The most likely cleavage site for a NOV56 peptide is between amino acids 22 and 23.

Table 56B. Encoded NOV56 protein sequence (SEQ ID NO:130).

MGLGGAAAALVAAAAAAAAAAAVVAGPRRRRRGCGARCHGPGRAAGKKMSKPRALEAAA
AAAATAPGLEMVERRGPGRPRTDGESVFTGQSKIYSYSPNKC SGMRFP LQEENS VTHHE
VKCQ GKPLAGIYRKREEKRNTGNAVQSAMKSKKQKIKDARRGP LQGKTQONHKL TDFYFPV
RRRSRKSAELQSEERKRIDELIESGKEEGMKIDLIDGKGRGVIATKHFSRGAFVVEYHG
DLIEITDAKKREALYAQDPSTGCMYYFYQLSKTYCVDATRETNRPGRPINH SKCGNCQT
KLHDIDGVPHLILIASQDIAAGEELLYDYGDRSKASIEAHPWLKH

A search of sequence databases reveals that the NOV56 amino acid sequence has 314
of 345 amino acid residues (91%) identical to, and 322 of 345 amino acid residues (93%)
similar to, the 345 amino acid residue ptnr:SPTREMBL-ACC:Q9NQR1 protein from Homo
sapiens (Human) (PR/SET DOMAIN CONTAINING PROTEIN 07). Public amino acid
databases include the GenBank databases, SwissProt, PDB and PIR.

NOV56 is expressed in at least Bone Marrow, Brain, Cervix, Dermis, Heart, Kidney,
Liver, Lung, Lymph node, Pancreas, Parietal Lobe, Pituitary Gland, Placenta, Skin, Spinal
Chord, Spleen, Thymus, Thyroid, Umbilical Vein, Adrenal Gland/Suprarenal gland, Aorta,
Ascending Colon, Brain, Buccal mucosa, Cartilage, Cervix, Chorionic Villus, Colon,
Coronary Artery, Duodenum, Heart, Kidney, Liver, Lung, Ovary, Parietal Lobe, Parotid
Salivary glands, Peripheral Blood, Prostate, Retina, Salivary Glands, Small Intestine,
Synovium/Synovial membrane, Testis, Tonsils, Umbilical Vein, and Urinary Bladder.
Expression information was derived from the tissue sources of the sequences that were
included in the derivation of the sequence of CG57348_01. The sequence is predicted to be
expressed in the following tissues because of the expression pattern of (GENBANK-ID:
gb:GENBANK-ID:AF287261|acc:AF287261.1) a closely related Homo sapiens PR/SET
domain containing protein 07 (SET07) mRNA, complete cds homolog in species Homo
sapiens :Bone Marrow, Brain, Cervix, Dermis, Heart, Kidney, Liver, Lung, Lymph node,
Pancreas, Parietal Lobe, Pituitary Gland, Skin, and Colon. This information was derived by
determining the tissue sources of the sequences that were included in the invention including
but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE
sources.

The disclosed NOV56 polypeptide has homology to the amino acid sequences shown
in the BLASTP data listed in Table 56C.

Table 56C. BLAST results for NOV56					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 15295688 ref XP 004668.4 (XM_004668	similar to PR/SET domain containing protein 07 (H. sapiens) [Homo sapiens]	403	306/359 (85%)	306/359 (85%)	e-149
gi 15042945 ref NP 065115.2 (NM_020382)	PR/SET domain containing protein 07 [Homo sapiens]	393	319/393 (81%)	324/393 (82%)	e-139
gi 15303323 ref XP 017300.2 (XM_017300)	PR/SET domain containing protein 07 [Homo sapiens]	240	145/165 (87%)	150/165 (90%)	3e-46
gi 7299871 gb AAF55 047.1 (AE003704)	CG3307 gene product [Drosophila melanogaster]	689	100/230 (43%)	143/230 (61%)	2e-45
gi 17554790 ref NP 498417.1 (NM_066016)	T26A5.7.p [Caenorhabditis elegans]	242	83/203 (40%)	125/203 (60%)	2e-33

Table 56D lists the domain descriptions from DOMAIN analysis results against NOV56. This indicates that the NOV56 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 56D. Domain Analysis of NOV56
gnl Smart smart00317, SET, SET (Su(var)3-9, Enhancer-of-zeste, Trithorax) domain; Putative methyl transferase, based on outlier plant homologues
CD-Length = 125 residues, 96.8% aligned
Score = 107 bits (266), Expect = 1e-24

Association of SET domain and myotubularin-related proteins modulates growth control. The PR domain of the Rb-binding zinc finger protein RIZ1 is a protein binding interface and is related to the SET domain functioning in chromatin-mediated gene expression.

10

SET domains appear to be protein-protein interaction domains. It has been demonstrated that SET domains mediate interactions with a family of proteins that display similarity with dual-specificity phosphatases (dsPTPases) [2]. A subset of SET domains have been called PR domains. These domains are divergent in sequence from other SET domains, but also appear to mediate protein-protein interaction [3].

15

The SET domain is a highly conserved, approximately 150-amino acid motif implicated in the modulation of chromatin structure. It was originally identified as part of a

larger conserved region present in the *Drosophila* Trithorax protein and was subsequently identified in the *Drosophila* Su(var)3-9 and 'Enhancer of zeste' proteins, from which the acronym SET is derived. Studies have suggested that the SET domain may be a signature of proteins that modulate transcriptionally active or repressed chromatin states through chromatin remodeling activities.

By sequencing cDNAs randomly selected from a cDNA library derived from a human immature myeloid cell line, Nomura et al. (1994) isolated a cDNA encoding SETDB1, which they called KIAA0067. The deduced SETDB1 protein has 1,291 amino acids. Northern blot analysis detected SETDB1 expression in all 16 human tissues examined.

In the course of searching sequence databases for proteins containing SET domains, Harte et al. (1999) identified the SETDB1 sequence. They determined that SETDB1 has a C-terminal SET domain that is well-conserved except that it contains a 347-amino acid insertion between its most highly conserved regions. The authors found that the *C. elegans* YNCA gene product is highly similar to SETDB1 and also contains a bifurcated SET domain.

Nomura et al. (1994) mapped the SETDB1 gene to chromosome 1 using a somatic cell hybrid mapping panel. By FISH and radiation hybrid mapping, Harte et al. (1999) mapped the SETDB1 gene to 1q21.

The disclosed NOV56 nucleic acid of the invention encoding a PR_SET_domain protein-like protein includes the nucleic acid whose sequence is provided in Table 56A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 56A while still encoding a protein that maintains its PR_SET_domain protein-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 5 percent of the bases may be so changed.

The disclosed NOV56 protein of the invention includes the PR_SET_domain protein-like protein whose sequence is provided in Table 56B. The invention also includes a mutant

or variant protein any of whose residues may be changed from the corresponding residue shown in Table 56B while still encoding a protein that maintains its PR_SET_domain protein-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 9 percent of the residues may be so changed.

5 The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

 The above defined information for this invention suggests that this PR_SET_domain protein-like protein (NOV56) may function as a member of a “PR_SET_domain protein family”. Therefore, the NOV56 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

 The NOV56 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the PR_SET_domain protein-like protein (NOV56) may be useful in gene therapy, and the PR_SET_domain protein-like protein (NOV56) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from CNS disorders, brain disorders including epilepsy, eating disorders, schizophrenia, ADD; cancer; heart disease; inflammation and autoimmune disorders including Crohn’s disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, blood disorders; psoriasis colon cancer, leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, pancreatic disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer, or other pathologies or conditions. The NOV56 nucleic acid encoding the PR_SET_domain protein-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

 NOV56 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV56 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods

known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV56 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV57

A disclosed NOV57 nucleic acid of 5896 nucleotides (also referred to as CG57650-01) encoding a non-muscle myosin heavy chain B-like protein is shown in Table 57A. The start and stop codons are in bold letters.

Table 57A. NOV57 nucleotide sequence (SEQ ID NO:131).

GGCGGTGCTGCGGGACGAAGGCGAGGAGGAGGCGGAGGTGGAGCTGGCGGAGAGCGGGAG
GCGGCTGCGACTGCCGCGGGACCAGATCCAGCGCATGAACCCGCGCCAAGTTCAGCAAGG
CCGAGGACATGGCCGAGCTGACCTGCCCTCAACGAGGCCTCGGTCTGCACAACCTCCGGG
AGCGGTACTACTCCGGCCTCATCTACACGTACTCCGGCCTTTCTGTGTGGTTCATCAACC
CGTACAAGCAGCTTCCCATCTACACAGAAGCCATTGTGGAGATGTACCGGGGCAAGAAGC
GCCACGAGGTGCCACCCACGTGTACGCAGTGACCGAGGGGGCCTATCGGAGCATGCTGC
AGGATCGTGAGGACCAGTCCATTCTCTGCACGGGAGAGTCTGGAGCTGGGAAGACGGAAA
ACACCAAGAAGGTCATCCAGTACCTCGCCACGTGGCGTCGTCTCCAAAGGGCAGGAAGG
AGCCGGGTGTCCCCGGTGAGCTGGAGCGGCAGCTGCTTCAGGCCAACCCCATCTAGAGG
CCTTTGGCAATGCCAAGACAGTGAAGAATGACAACTCCTCCCGATTGGCAAATTCATCC
GCATCAACTTTGATGTTGCCGGGTACATCGTGGGCGCCAACATTGAGACCTGTCTGTCTGG
AGAAGTCGCGGGCCATCCGCCAGGCCAAGGACGAGTGACGCTTCCACATCTTCTACCAGC
TGCTGGGGGGCGCTGGAGAGCATGGCTGCCGAGAACTCCTCCTCGAGCCCTGTCTCCACT
ACCGGTTCTTGACCAACGGGCCGTCTCTCCTCGGCCAGGAGCGGGAACCTTCTCCAGG
AGACGCTGGAGTCGCTGCGGGTCTTGGGATTACGCCACGAGGAAATCATCTCCATGCTGC
GGATGGTCTCAGCAGTTCTCCAGTTTGGCAACATTGCCTTGAAGAGAGAACCGAACACCG
ATCAAGCCACCATGCCTGACAACACAGCTGCACAGAAGCTCTGCCGCCTCTTGGGACTGG
GGGTGACGGATTTCTCCCGAGCCTTGCTCACCCCTCGCATCAAAGTTGGCCGAGACTATG
TGCAGAAAGCCCAGACTAAGGAACAGGCTGACTTCGCGCTGGAGGCCCTGGCCAAGGCCA
CCTACGAGCGCCTCTTCCGCTGGCTGGTTCTGCGCCTCAACCGGGCCTTGGACCGCAGCC
CCCGCCAAGGCGCCTCCTTCTGGGCATCCTGGACATCGCGGGCTTTGAGATCTTCCAGC
TGAACCTCTTCGAGCAGCTCTGCATCAACTACACCAACGAGAAGCTGCAGCAGCTCTTCA
ACCACACCATGTTTCGTGCTGGAGCAGGAGGAGTACCAGCGTGAGGGCATCCCCTGGACCT
TCCTCGACTTTGGCCTCGACCTGCAGCCCTGCATCGACCTCATCGAGCGGCCGCGCCAACC
CCCCTGGACTCCTGGCCCTGCTGGATGAGGAGTGCTGGTTCCCGAAGGCCACAGACAAGT
CGTTTGTGGAGAAGGTAGCCCAGGAGCAGGGCGGCCACCCCAAGTTCAGCGGCCGAGGC
ACCTGCGGGATCAGGCCGACTTCAGTGTCTTCCACTACGCGGGCAAGGTCGACTACAAGG
CCAACGAGTGGCTGATGAAAAACATGGACCCTCTGAATGACAACGTTCGACGCTTGTCTCC
ACCAGAGCACAGACCGGCTGACGGCAGAGATCTGGAAAGACGTGGAGGGCATCGTGGGGC
TGGAAACAGGTGAGCAGCTTGGGCGACGGCCACCAGGTGGCCGCCCCCGTGGGGTATGT
TCCGGACAGTGGGACAGCTCTACAAGGAGTCCCTGAGCCGCCCTCATGGCCACATCAGCA
ACACCAACCCAGTTTTGTCCGCTGCATTGTCCCCAACCCAGAGAAGAGGGTGGGAAGC
TGGAGCCGCGGCTGGTGTGGACCAGCTTCGCTGCAACGGGGTCTGGAGGGCATCCGCA
TCTGTGCCAGGGCTTCCCCAACCGCATCTCTTCCAGGAGTTCGGCAGCGATACGAGA
TCCTGACACCCAATGCCATCCCCAAGGGCTTCATGGATGGGAAGCAGGCCTGTGAAAAGA
TGATCCAGGCGCTGGAACCTGGACCCCAACCTCTACCGCGTGGGACAGAGCAAGATCTTCT
TCCGGGCTGGGGTCTTGGCCAGCTGGAAGAGGAGCGAGACCTGAAGGTCACCGACATCA
TCGTCTCTTCCAGGCAGCTGCCCGGGGATACCTGGCTCGCAGGGCCTTCCAGAAGCGCC

AGCAGCAGCAGAGCGCCCTGAGGGTGATGCAGCGGAAC TGCGCGGCCTACCTCAAGCTGA
 GACACTGGCAGTGGTGGCGGCTGTTTACCAAGGTGAAGCCACTGCTGCAGGTGACGCGGC
 AGGATGAGGTGCTGCAGGCACGGGCCAGGAGCTGCAGAAAGTGCAGGAGCTACAGCAGC
 AGAGCGCCCGCAAGTTGGGGAGCTCCAGGGCCGAGTGACACAGCTGGAAGAGGAGCGCG
 CCCGCTGGCAGAGCAATTGCGAGCAGAGGCAGAACTGTGTGCAGAGGCCGAGGAGACGC
 GGGGGAGGCTGGCAGCCCGCAAGCAGGAGCTGGAGCTGGTGGTGTGAGAGCTGGAGGCTC
 GCGTGGGCGAGGAGGAGGAGTGCAGCCGTCAAATGCAAACCGAGAAGAAGAGGCTACAGC
 AGCACATACAGGAGCTAGAGGCCACCTTGAGGCTGAGGAGGCTGCGCGGCAGAGCTGC
 AGCTGGAGAAGGTGACGACAGAGGCAGAAATGAAGAAATTTGAAGAGGACCTGCTGCTCC
 TGGAAAGACCAGAATTCCAAGCTGAGCAAGGAACTGCTGGAAGATCGTCTGGCCGAGTTCT
 CATCCCAGGCAGCTGAGGAGGAGGAGAAGGTCAAGAGCCTCAATAAGCTACGGCTCAAAT
 ATGAGGCCACAATCGCAGACATGGAGGACCGCCTACGGAAGGAGGAGAAGGTCGCCAGG
 AGCTGGAGAAGCTGAAGCGGAGGCTGGATGGGGAGAGCTCAGAGCTGCAGGAGCAGATGG
 TGGAGCAGCAACAGCGGGCAGAGGAGCTGCGGGCCAGCTGGGCGGGAAGGAGGAGGAGC
 TGCAGGCTGCCCTGGCCAGGGCAGAAGACGAGGCTGGGGCCCGGGCCAGCTGCTGAAAT
 CCCTGCGGGAGGCTCAAGCAGCCCTGGCCGAGGCCGAGGCCAGGAGGACCTGGAGTCTG
 AGCGTGTGGCCAGGACCAAGGCGGAGAAGCAGCGCCGGGACCTGGGCGAGGAGCTGGAGG
 CGCTGCGGGGCGAGCTGGAGGACACGCTGGACTCCACCAACGCACAGCAGGAGCTCAGGT
 CCAAGAGGGAACAGGAGGTGACGGAGCTGAAGAAGACTCTGGAGGAGGAGACTCGCATCC
 ACGAGGCGGCAGTGACGAGCTGAGGCAGCGCCACGGCCAGGCCCTGGGGGAGCTGGCGG
 AGCAGCTGGAGCAGGCCCGGAGGAAAGGTGCATGGGAGAAGACCCGGCTGGCCCTGGAGG
 CCGAGGTGTCCGAGCTGCGGGCAGAACTGAGCAGCCTGCAGACTGCACGTCAGGAGGGTG
 AGCAGCGGAGGCGCCGCTGGAGTTACAGCTGCAGGAGGTGCAGGGCCGGGCTGGTGATG
 GGGAGAGGGCACGAGCGGAGGCTGCTGAGAAGGTCCCTTCCCTGCAGGCTGAACTGGAGA
 ATGTGTCTGGGGCGCTGAACGAGGCTGAGTCCAAACCATCCGTCTTAGCAAGGAGCTGA
 GCAGCACAGAAGCCAGCTGCACGATGCCAGGAGCTGCTGCAGGAGGAGACCAGGGCGA
 AATTGGCCCTTGGGGTCCCGGGTGCAGCCATGGAGGCTGAGGCAGCCGGGCTGCGTGAGC
 AGCTGGAGGAGGAGGAGCTGCCAGGGAACGGGCGGACCACCAACCACCTCTCTCTCTCT
 CCCCTCAGCTTTCCGAGTGGCGGGCGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
 GGGAGGAGGACGCGCGCCGGGAGCCCGGGAGGCCGAGGCCCTGACCCAGCGCCTGGCAG
 AAAAGACAGAGACCGTGGATCGGCTGGAGCGGGGCCCGCCCGGCTGGGGCAGGAGCTGG
 ACGACGCCACCATGGACCTGGAGCAGCAGCGGAGCTTGTGAGCACCTGGAGAAGAAGC
 AGCGCAAGTTTGACCAGCTTCTGGCAGAGGAGAAGGCAGCTGTACTTCGGGCACTGGAGG
 AACGTGAGCGGGCCGAGGCAGAGGGCCGGGAGCGTGAGGCTCGGGCCCTGTCACTGACAC
 GGGCACTGGAGGAGGAGCAGGAGGCACGTGAGGAGCTGGAGCGGCAGAACCGGGCCCTGC
 GGGCTGAGCTGGAGGCACTGCTGAGCAGCAAGGATGACGTGCGCAAGAGCGTGATGAGC
 TGGAAACGAGCCTGCCGGGTAGCAGAACAGGCAGCCAATGATCTGCGAGCACAGGTGACAG
 AACTGGAGGATGAGCTGACAGCGGCCGAGGATGCCAAGCTGCGTCTGGAGGTGACTGTGC
 AGGCTCTCAAGACTCAGACTGAGCGTGACCTGCAGGGCCGTGATGAGGCTGGTGAAGAGA
 GGGCGGAGCTGGCCAAGCAGCTGAGAGATGCAGAGGTGGAGCGGGATGAGGAGCGGA
 AGCAGCGCACTCTGGCCGTGGCTGCCCCGCAAGAAGCTGGAGGGAGAGCTGGAGGAGCTGA
 AGGCTCAGATGGCCTCTGCCGGCCAGGGCAAGGAGGAGGCGGTGAAGCAGCTTCGCAAGA
 TGCAGGCCCAGATGAAGGAGCTATGGCGGGAGGTGGAGGAGACACGCACCTCCCGGGAGG
 AGATCTTCTCCAGAATCGGGAAAGTGAAAAGCGCCTCAAGGGCCTGGAGGCTGAGGTGC
 TGCGGCTGCAGGAGGAACTGGCCGCTCGGACCGTGCTCGGCGGCAGGCCCAGCAGGACC
 GGGATGAGATGGCAGATGAGGTGGCCAATGGTAACCTTAGCAAGGCAGCCATTCTGGAGG
 AGAAGCGTCAGCTGGAGGGCGCCTGGGGCAGTTGGAGGAAGAGCTGGAGGAGGAGCAGA
 CAACTCAGAGCTGCTCAATGACCGCTACCGCAAGCTGCTCCTGCAGGTAGAGTCACTGAC
 CACAGAGCTGTGAGCTGAGCGCAGTTTCTCAGCCAAGGCAGAGAGCGGGCGGCAGCAGCT
 GGAACGGCAGATCCAGGAGCTACGGGGACGCCCTGGGTGAGGAGGATGCTGGGGCCCGTGC
 CCGCCACAAGATGACCATTGCTGCCCTTGAGTCTAAGTTGGGCCAGGCTGAGGAGCAGCT
 AGAGCAAGAGACCAGAGAGCGCATCTCTCTGGAAGCTGGTGCCCAAAAGTTAAGAAGC
 GGCTTAAAGAGGTGGTGCTCCAGGTGGAGGAGGAGCGGAGGGTGGCTGACCAGCTCCGGG
 ACCAGCTGGAGAAGGGAAACCTTCGAGTCAAGCAGCTGAAGCGGCAGCTGGAGGAGGCCG
 AGGAGGAGGCATCCCGGGCTCAGGCCGGCCGAGGCTGCAGCGTGAGCTGGAAGATG
 TCACAGAGTGGGCCGAGTCCATGAACCGTGAAGTGACCACACTGAGGAACCGGCTTCGAC
 GCGGCCCCCTCACCTTCACCANCCGCACGGTGCGCCAGGTCTTCCGACTAGAGGAGGGCG
 TGGCATCCGACGAGGAGGCAGAGGAAGCACAGCCTGGGTCTGGGCCATCCCCTCTCACTC
 CTGCTGCTGCCATGCTCTGCCCTCCCTTCTGGTTGCTCTGAGGGTTCGAGCTTCCCTC
 TGGGACTAAAGGAGTGTCTTTACCCTCCAGCCTCCCGGCTTGGCAGAAATAAACTCC
 AACCCGAATGGAAAAA

In a search of public sequence databases, the NOV57 nucleic acid sequence, located on chromosome 10 has 3509 of 4934 bases (71%) identical to a gb:GENBANK-ID:RABMHCP|acc:M77812.1 mRNA from *Oryctolagus cuniculus* (Rabbit myosin heavy chain mRNA, complete cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV57 polypeptide (SEQ ID NO:132) encoded by SEQ ID NO:131 has 1673 amino acid residues and is presented in Table 57B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV57 has no signal peptide and is likely to be localized in the nucleus with a certainty of 0.9600.

Table 57B. Encoded NOV57 protein sequence (SEQ ID NO:132).

MAELTCLNEASVLHNLRRERYYSGLIYTYSGFLFCVVINPYKQLPIYTEAIVEMYRGKKRHE
VPPHVVAVTEGAYRSMQDREDQSILCTGESGAGKTENTKKVIQYLAHVASSPKGRKEPG
VPGELERQLLQANPILEAFGNAKTVKNDNSSRFGKFIRINFDVAGYIVGANIETCALLEKS
RAIRQAKDECSFHIIFYQLLGGAGEHGCRELLLEPCSHYRFLTNGPSSSPGQERELFQETL
ESLRVLGFSHEEIIISMLRMVSAVLQFGNIALKRERNTDQATMPDNTAAQKLCRLGLGVT
DFSRAALLTPRIKVGGRDYVQKAQTKEQADFALEALAKATYERLFRWLVLRLNLRALDRSPRQ
GASFLGILDIAQFEIFQLNSFEQLCINYTNEKLQQLFNHTMFVLEQEEYQREGIPWTFDL
FGLDLQPCIDLIERPANPPGLLALLDEECWFPPKATDKSFVEKVAQEQQGHPKQRPRLR
DQADFSVLHYAGKVVDYKANEWLMKNMDPLNDNVAALLHQSTDRLTAEIWKDVEGIVGLEQ
VSSLGDGPPGGRPRRGMFRTVGQLYKESLSRLMATLSNTNPSFVRCIVPNHEKRVGKLEP
RLVLDQLRCNGVLEGIRICRQGFPNRIILFQEFRQRYEILTPNAIPKGFMDGKQACEKMIQ
ALELDPNLYRVGQSKIIFFRAGVLAQLEEDDLKVTDIIVSFQAAARGYLARRAFQKRQQQ
QSALRVMQRNCAAYLKLRLHWQWRLFTKVKPLLQVTRQDEVLQARAQELQKVQELQQQSA
REVGELQGRVTQLEEEERARLAEQLRAEAELCAEAETRGRLAARKQELELVSELEARVG
EEEECSRQMOTTEKKRLQHQHIELEAHLEAEEGARQKLQLEKVTTEAKMKKFEEDLLLLLED
QNSKLSKELLEDRLAEFSSQAEEEEKVKSLNKLRLKYEATIAMEDRLRKEEKGRQLE
KLKRRLDGESSELQEQMVEQQQRAEELRAQLGRKEEELQAALARAEDEGGARAQLLKSRL
EAQAALAEAEAEQEDLESERVARTKAQKRRDLGEELEALRGELEDTLDSNAQQELRSKR
EQEVTELKKTLEETRIHEAAVQELRQRHGQALGELAEQLEQARRKGAWEKTRLALEAEV
SELRAELSSLQTARQEGEQRRRLLELQLQEVQGRAGDGERARAEAEKVPSLQAELENV
GALNEAESKTIIRLSKELSSTEAQLHDAQELLQEETRAKLALGSRVRAMEAEAGLREQLE
EEAAARERADHQPPSLSSPOLSEWRRRQEEAGALEAGEEARRRAAREAEALTQRLAEKT
ETVDRLERGRRRLGQELDDATMDLEQQRQLVSTLEKKQRKFDQLLAEEKAQVLRVEERE
RAEAEGREREARALSLTRALEEEQEAEEELERQNRALRAELEALLSSKDDVGKSVHELER
ACRVAEQAAANDLRAQVTELEDELTAEDAKLRLEVTVQALKTQHERDLQGRDEAGEERRR
QLAKQLRDAEVERDEERKQRTLAVAARKKLEGELEELKAQMASAGQGKEEAVKQLRKMQA
QMKELWREVEETRTSREEIFSQNRESEKRLKGLEAEVLRLQEELAASDRARRQAQQRDE
MADEVANGNLSKAAILEEKRLQLEGRGLQLEEELEEEQTTQSCSMTATASCSCR

A search of sequence databases reveals that the NOV57 amino acid sequence has 1149 of 1661 amino acid residues (69%) identical to, and 1395 of 1661 amino acid residues (83%) similar to, the 1976 amino acid residue ptnr:SWISSNEW-ACC:P35580 protein from *Homo sapiens* (Human) (MYOSIN HEAVY CHAIN, NONMUSCLE TYPE B (CELLULAR MYOSIN HEAVY CHAIN, TYPE B) (NMMHC-B)). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV57 is expressed in at least adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea and uterus. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV57 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 57C .

Table 57C. BLAST results for NOV57					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 13928704 ref NP_113708.1 (NM_031520)	myosin heavy chain 11 [Rattus norvegicus]	1976	1144/1664 (68%)	1389/1664 (82%)	0.0
gi 212449 gb AAA48985.1 (M93676)	nonmuscle myosin heavy chain [Gallus gallus]	1976	1148/1664 (68%)	1395/1664 (82%)	0.0
gi 1346640 sp P35580 MYHA HUMAN	Myosin heavy chain, nonmuscle type B (Cellular myosin heavy chain, type B) (Nonmuscle myosin heavy chain-B) (NMMHC-B)	1976	1149/1664 (69%)	1395/1664 (83%)	0.0
gi 212451 gb AAA48987.1 (M93676)	nonmuscle myosin heavy chain [Gallus gallus]	1997	1148/1685 (68%)	1395/1685 (82%)	0.0
gi 212450 gb AAA48986.1 (M93676)	nonmuscle myosin heavy chain [Gallus gallus]	1986	1148/1674 (68%)	1395/1674 (82%)	0.0

Tables 57D-H list the domain descriptions from DOMAIN analysis results against NOV57. This indicates that the NOV57 sequence has properties similar to those of other proteins known to contain this domain.

Table 57D. Domain Analysis of NOV57
gnl Pfam pfam00063 , myosin_head, Myosin head (motor domain).
CD-Length = 670 residues, 99.6% aligned
Score = 922 bits (2384), Expect = 0.0

gnl|Smart|smart00242, MYSc, Myosin. Large ATPases.; ATPase; molecular motor. Muscle contraction consists of a cyclical interaction between myosin and actin. The core of the myosin structure is similar in fold to that of kinesin.

Score = 866 bits (2238), Expect = 0.0

[gnl|Pfam|pfam01576](#), Myosin_tail, Myosin tail. The myosin molecule is a multi-subunit complex made up of two heavy chains and four light chains it is a fundamental contractile protein found in all eukaryote cell types. This family consists of the coiled-coil myosin heavy chain tail region. The coiled-coil is composed of the tail from two molecules of myosin. These can then assemble into the macromolecular thick filament. The coiled-coil region provides the structural backbone the thick filament.

Score = 344 bits (882), Expect = 3e-95

[gnl|Pfam|pfam01496](#), V_ATPase_sub_a, V-type ATPase 116kDa subunit family. This family consists of the 116kDa V-type ATPase (vacuolar (H⁺)-ATPases) subunits, as well as V-type ATP synthase subunit i. The V-type ATPases family are proton pumps that acidify intracellular compartments in eukaryotic cells for example yeast central vacuoles, clathrin-coated and synaptic vesicles. They have important roles in membrane trafficking processes. The 116kDa subunit (subunit a) in the V-type ATPase is part of the V0 functional domain responsible for proton transport. The a subunit is a transmembrane glycoprotein with multiple putative transmembrane helices it has a hydrophilic amino terminal and a hydrophobic carboxy terminal. It has roles in proton transport and assembly of the V-type ATPase complex. This subunit is encoded by two homologous gene in yeast VPH1 and STV1.

Score = 57.0 bits (136), Expect = 8e-09

gnl|Pfam|pfam00769, ERM, Ezrin/radixin/moesin family. This family of proteins contain a band 4.1 domain (pfam00373), at their amino terminus. This family represents the rest of these proteins.

Score = 52.0 bits (123), Expect = 3e-07

Phylogenetic analysis currently places myosins into 15 classes. The conventional myosins

which form filaments in muscle and non-muscle cells form class II. There has been extensive characterization of these myosins and much is known about their function. With the exception of class I and class V myosins, little is known about the structure, enzymatic properties, intracellular localization and physiology of most unconventional myosin classes. (See Sellers JR, 2000, *Biochim Biophys Acta* 1496:3-22). The discovery in of a huge diversity within the myosin superfamily has been coupled with an understanding of the role of these motor proteins in various cellular functions. Extensive studies have revealed that myosin isoforms are not only involved in muscle contraction but also in crucial functions of many specialized mammalian cells such as melanocytes, kidney and intestinal brush border microvilli, nerve growth cones or inner ear hair cells. A search for genes involved in the pathology of human genetic deafness resulted in identification of three novel myosins: myosin VI, myosin VIIA and, very recently, myosin XV. Recently, mutations have been detected within these genes that have been found to affect the hearing process (See Redowicz MJ, 1999, *J Muscle Res Cell Motil* 20:241-8). Class II non-muscle myosins are implicated in diverse biological processes such as cytokinesis, cellularization, cell shape changes and gastrulation. Two distinct non-muscle myosin heavy chain genes have been reported in all vertebrates: non-muscle myosin heavy chain-A (NMHC-A) and -B (NMHC-B). Whole mount in situ hybridization with tailbud stage embryos of *Xenopus* showed that NMHC-A mRNA is predominantly expressed in the epidermis, whereas NMHC-B mRNA is expressed in the somites, brain, eyes and branchial arches. Interestingly, the expression of NMHC-B in developing somites was gradually restricted to the center of each somite as differentiation proceeds. DAPI nuclear staining demonstrated that NMHC-B mRNA is colocalized with the nuclei or perinuclear area. In animal cap experiments, treatment with activin A or ectopic expression of Xbra and an activated form of Xlim1 markedly up-regulates NMHC-B as well as muscle actin mRNAs and slightly down-regulates NMHC-A mRNA, consistent with NMHC-B expression in the somitic muscle and NMHC-A expression in the epidermis. (See Bhatia et al., 1998, *Mech Dev* 78:33-6).

The disclosed NOV57 nucleic acid of the invention encoding a non-muscle myosin heavy chain B-like protein includes the nucleic acid whose sequence is provided in Table 57A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 57A while still encoding a protein that maintains its non-muscle myosin heavy chain B-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic

acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 29 percent of the bases may be so changed.

The disclosed NOV57 protein of the invention includes the non-muscle myosin heavy chain B-like protein whose sequence is provided in Table 57B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 57B while still encoding a protein that maintains its non-muscle myosin heavy chain B-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 31 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this non-muscle myosin heavy chain B-like protein (NOV57) may function as a member of a “non-muscle myosin heavy chain B family”. Therefore, the NOV57 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV57 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the non-muscle myosin heavy chain B-like protein (NOV57) may be useful in gene therapy, and the non-muscle myosin heavy chain B-like protein (NOV57) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will

have efficacy for treatment of patients suffering from hypertension and vasospasm of the coronary and cerebral arteries, coronary artery spasm, arteriosclerosis, hypertrophic cardiomyopathy, inflammatory diseases such as asthma, cancer, or other pathologies or conditions. The NOV57 nucleic acid encoding the non-muscle myosin heavy chain B-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV57 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV57 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV57 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV58

A disclosed NOV58 nucleic acid of 688 nucleotides (also referred to as CG57766-01) encoding a plasma retinol binding protein-like protein is shown in Table 58A. The start and stop codons are in bold letters.

Table 58A. NOV58 nucleotide sequence (SEQ ID NO:133).
<p>CCCAACCACGGCCAGGCTTGCGCGCGGTTCCCTCCCGGTGGGCGGATTCTGGGCAAGATGAAGTGGGT GTGGGCGCTCTTGCTGTTGGCGGCGCTGGGCAGCGGCCGCGCGGAGCGCGACTGCCGAGTGAGCAGCTTC CGAGTCAAGGAGAACTTCGACAAGGCTCGCTTCTCTGGGACCTGGTACGCCATGGCCAAGAAGGACCCCG AGGGCCTCTTTCTGCAGGACAACATCGTCGCGGAGTTCTCCGTGGACGAGACCGGCCAGATGAGCGCCAC AGCCAAGGGCCGAGTCCGTCTTTTGAATAACTGGGACGTGTGCGCAGACATGGTGGGCACCTTCACAGAC ACCGAGGACCTGCCAAGTTCAAGATGAAGTACTGGGGCGTAGCCTCCTTTCTCCAGAAAGGAAATGATG ACCACTGGATCGTCGACACAGACTACGACACGTATGCCGTGCAGTACTCCTGCCGCCTCCTGAACCTCGA TGGCACCTGTGCTGACAGCTACTCCTTCGTGTTTTCCCGGGACCCCAACGGCCTGCCCCCAGAAGCGCAG AAGATTGTAAGGCAGCGGCAGGAGGAGCTGTGCCTGGCCAGGCAGTACAGGCTGATCGTCCACAACGGTT ACTGCGATGGCAGATCAGAAAGAAACCTTTTGTAGCAAGGGCGAATTCAGCACACTG</p>

In a search of public sequence databases, the NOV58 nucleic acid sequence, located on chromosome 10q23-24 has 657 of 673 bases (97%) identical to a gb:GENBANK-ID:HSRBP1|acc:X00129.1 mRNA from Homo sapiens (Human mRNA for retinol binding protein (RBP)). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV58 polypeptide (SEQ ID NO:134) encoded by SEQ ID NO:133 has 201 amino acid residues and is presented in Table 58B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV58 has a signal peptide and is likely to be localized extracellularly with a certainty of 0.3700. The most likely cleavage site for a NOV58 peptide is between amino acids 18 and 19.

Table 58B. Encoded NOV58 protein sequence (SEQ ID NO:134).	
MKWVWALLLLAALGSGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIV AEFSVDETQMSATAKGRVRLNNDVDCADMGTFDTEDPAKFKMKYWGVSFLQKGND DHWIVDTDYDTYAVQYSCRLNLDGTCADSYFVSFVRDPLPPEAQKIVRQRQEELCLA RQYRLIVHNGYCDGRSERNLL	

A search of sequence databases reveals that the NOV58 amino acid sequence has 196 of 201 amino acid residues (97%) identical to, and 197 of 201 amino acid residues (98%) similar to, the 199 amino acid residue ptnr:SWISSNEW-ACC:P02753 protein from Homo sapiens (Human) (PLASMA RETINOL-BINDING PROTEIN PRECURSOR (PRBP) (RBP)). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV58 is expressed in at least adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea and uterus. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV58 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 58C.

Table 58C. BLAST results for NOV58					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 18088326 gb AAH20633.1 AAH20633 (BC020633)	Similar to retinol binding protein 4, plasma [Homo sapiens]	201	200/201 (99%)	200/201 (99%)	e-113
gi 2136468 pir I46257	retinol binding protein precursor - horse	201	186/201 (92%)	195/201 (96%)	e-107
gi 3041715 sp P27485 RETB_PIG	PLASMA RETINOL- BINDING PROTEIN PRECURSOR (PRBP) (RBP)	201	184/201 (91%)	193/201 (95%)	e-106

gi 1710096 sp P06912 RETB RABIT	PLASMA RETINOL-BINDING PROTEIN PRECURSOR (PRBP) (RBP)	201	183/201 (91%)	194/201 (96%)	e-106
gi 89271 pir A39486	plasma retinol-binding protein precursor - pig	201	184/201 (91%)	192/201 (94%)	e-106

Table 58D lists the domain descriptions from DOMAIN analysis results against NOV58. This indicates that the NOV58 sequence has properties similar to those of other proteins known to contain this domain.

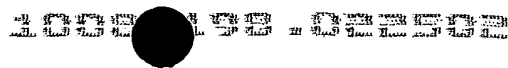
Table 58D. Domain Analysis of NOV58

gnl|Pfam|pfam00061, lipocalin, Lipocalin / cytosolic fatty-acid binding protein family. Lipocalins are transporters for small hydrophobic molecules, such as lipids, steroid hormones, bilins, and retinoids. Alignment subsumes both the lipocalin and fatty acid binding protein signatures from PROSITE. This is supported on structural and functional grounds. Structure is an eight-stranded beta barrel.

CD-Length = 145 residues, 100.0% aligned

Score = 102 bits (255), Expect = 2e-23

Vitamin A is mobilized from liver stores and transported in plasma in the form of the lipid alcohol retinol, bound to a specific transport protein, retinol-binding protein (RBP). A great deal is known about the chemical structure, metabolism, and biological roles of RBP. RBP is a single polypeptide chain with molecular weight close to 20,000. RBP interacts strongly with plasma prealbumin, and normally circulates in plasma as a 1:1 molar RBP-prealbumin complex. Both the primary and the tertiary structure of prealbumin are known, and the primary structure of RBP has recently been reported. Much information is available about the protein-protein and protein-ligand interactions that are involved in this transport system. Many clinical studies have examined the effects of a variety of diseases on the plasma levels of RBP and prealbumin in humans. Plasma RBP levels are low in patients with liver disease and are high in patients with chronic renal disease. These findings reflect the facts that RBP is produced in the liver and mainly catabolized in the kidneys. Delivery of retinol to extra-hepatic tissues appears to involve specific cell surface receptors for RBP. Vitamin A mobilization from the liver, and delivery to peripheral tissues, is highly regulated by factors that control the rates of RBP production and secretion. Retinol deficiency specifically blocks the secretion of RBP, so that plasma RBP levels fall and liver RBP levels rise. Injection of retinol into vitamin A-deficient rats stimulates the rapid secretion of RBP from the liver into the plasma (See Goodman D.S., 1980, Ann N Y Acad Sci 348:378-90).



The disclosed NOV58 nucleic acid of the invention encoding a plasma retinol binding protein-like protein includes the nucleic acid whose sequence is provided in Table 58A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 58A while still encoding a protein that maintains its plasma retinol binding protein-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 3 percent of the bases may be so changed.

The disclosed NOV58 protein of the invention includes the plasma retinol binding protein-like protein whose sequence is provided in Table 58B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 58B while still encoding a protein that maintains its plasma retinol binding protein-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 3 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this plasma retinol binding protein-like protein (NOV58) may function as a member of a "plasma retinol binding protein family". Therefore, the NOV58 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV58 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the plasma retinol binding protein-like protein (NOV58) may be useful in gene therapy, and the plasma retinol binding protein-like protein (NOV58) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from cervical dysplasias and cancer, breast cancer, phenylketonuria, liver diseases, kidney diseases, alzheimers, infection and inflammations, or other pathologies or conditions. The NOV58 nucleic acid encoding the plasma retinol binding protein-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV58 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV58 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV58 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV59

A disclosed NOV59 nucleic acid of 1647 nucleotides (also referred to as CG57566-01) encoding a HIV-1 inducer of short transcripts binding protein like protein-like protein is shown in Table 59A. The start and stop codons are in bold letters.

Table 59A. NOV59 nucleotide sequence (SEQ ID NO:135).

ATGGCCGGCGGCGTGGACGGCCCCATCGGGATCCCGTTCCTCCCGACACAGCAGCGACATC
CTGAGTGGGCTGAACGAGCAGCGGACGCAGGGCCTGCTGTGCGACGTGGTGATCCTGGTG
GAGGGCCGCGAGTTCCCCACGCACCGCTCGGTGCTGGCCGCCTGCAGCCAGTACTTCAAG
AAGCTGTTCACGTCGGGCGCCGTGGTGGACCAGCAGAACGTGTACGAGATCGACTTCGTC
AGCGCCGAGGCGCTCACCGCGCTCATGGACTTCGCCTACACGCCACGCTCACCGTCAGC
ACAGCCAACGTGGGTGACATCCTCAGCGCGCCCGCCTGCTGGAGATCCCCGCCGTGAGC
CACGTGTGCGCCGACCTCCTGGACCGGCAGATCCTGGCGGCCGACGCGGGCGCCGACGCC
GGGCAGCTGGACCTTGTAGATCAAATTGATCAGCGCAACCTCCTCCGCGCCAAGGAGTAC
CTCGAGTTCTTCCAGAGCAACCCCATGAACAGCCTGCCCCCGCGGCCGCGCCGCGCT
GCCAGCTTCCCGTGGTCCGCCTTTGGGGCGTCCGATGATGACCTGGATGCCACCAAGGAG
GCCGTGGCCGCCGCTGTGGCCGCCGTGGCCGCGGGCGACTGCAACGGCTTAGACTTCTAT

GGGCCGGGCCCCCGGCCGAGCGGCCCCGACGGGGACGGGGACGAGGGCGACAGCAAC
 CCGGGTCTGTGGCCAGAGCGGGATGAGGACGCCCCACCGGGGTCTCTTTCCGCCGCCG
 GTGGCCCCGCCGGCCGACGACGAGCGGCGGAGAGGAGGAGGCC
 GCCTCGCTGTCTGGAGGCGGCCCCGAGCCGGGCGACTCTCCGGGCTTCCTGTCTGGGAGAC
 AGCGACGAGGAGTCTCGGGGCGGACGACAAGGGCGTTCATGGACTACTACCTGAAGTACTTC
 AGCGGCGCCACGACGGCGACGTCTACCGGCCTGGTCGAGAAGGTGGAGAAGAAGATC
 CGAGCCAAGGCCTTCCAGAAGTGCCCCATCTGCGAGAAGGTCTCCAGGGCGCCGGCAAG
 CTGCCGCGACACATCCGCACCCACACGGGCGAGAAGCCCTACGAGTGCAACATCTGCAAG
 GTCCGCTTACCAGGACAGGACAAGCTGAAGGTGCACATGCGGAAGCACACGGGCGAGAAG
 CCGTACCTGTGCCAGCAGTGCGGCGCCGCTTTGCCCACTACGACCTGAAGAACCAC
 ATGCGCGTGACACGGGCGTGCAGCCCTACAGTGCGACAGCTGCTGCAAGACCTTCGTC
 CGCTCCGACCACCTGCACAGACACCTCAAGAAAGACGGCTGCAACGGCGTCCCTCGCGC
 CGCGGCCGCAAGCCCCGCGTCCGGGGCGGGGCGCCCGACCCAGCCCGGGGGCCACCGCG
 ACCCCCGGCGCCCCCGCCAGCCAGCTCCCCGACGCCCGGCGCAACGGCCAGGAGAAG
 CACTTTAAGGACGAGGACGAGGACGAGGACGTGGCCAGCCCCGACGGCTTGGGCGCGTTG
 AATGTAGCGGGCGCCGGTGGAGGAGGTGACAGCGGAGGTGGCCCCGGGGCCGCCACCGAC
 GGTAACCTTACAGCCGGAAGTTCGCTTAA

In a search of public sequence databases, the NOV59 nucleic acid sequence has 1271 of 1560 bases (81%) identical to a gb:GENBANK-ID:AF097916|acc:AF097916.1 mRNA from Homo sapiens (Homo sapiens HIV-1 inducer of short transcripts binding protein (FBI1) mRNA, complete cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV59 polypeptide (SEQ ID NO:136) encoded by SEQ ID NO:135 has 548 amino acid residues and is presented in Table 59B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV59 has no signal peptide and is likely to be localized in the nucleus with a certainty of 0.8800.

Table 59B. Encoded NOV59 protein sequence (SEQ ID NO:136).

MAGGVDGPIGIPFPDHSSDILSGLNEQRTQGLLCDVVILVEGREFPTHRSLAACSQYFK
 KLFTSGAVVDQQNVYEIDFVSAEALTALMDFAYTATLTVSTANVGDLISAARLLEIPAVS
 HVCADLLDRQILAADAGADAGQLDLVDQIDQRLNLLRAKEYLEFFQSNPMNSLPPAAAAAA
 ASFPWSAFGASDDDLATKEAVAAVAAGDCNGLDFYGPAPPAERPPPTGDGDEGDSN
 PGLWPERDEDAPTGGFLFPPPVAPPAATQNGHYGRGEEEEASLSEAAPEPGDSPGFLSGD
 SDEESRADDKGVMDDYLYKYSFSAHDGDVYPAWSQKVEKKIRAKAFQKCPICEKVIQAGAK
 LPRHIRTHTGEKPYECNICKVRFTRQDKLKVMRKHTEKPYLCQQCGAAFAHNYDLKNH
 MRVHTGLRPYQCDSCCKTFVRSDDLHRLKKGDCNGVPSRRGRKPRVRGGAPDPSPGATA
 TPGAPAQPSPPDARRNGQEKHFKEDEDEDVASPDGLRLNVAGAGGGGDSGGGPGAATD
 GNFTAGLA

A search of sequence databases reveals that the NOV59 amino acid sequence has 344 of 458 amino acid residues (75%) identical to, and 362 of 458 amino acid residues (79%) similar to, the 584 amino acid residue ptnr:SPTREMBL-ACC:O95365 protein from Homo sapiens (Human) (HIV-1 INDUCER OF SHORT TRANSCRIPTS BINDING PROTEIN). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV59 is expressed in at least tumor, inflammed, and brain. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

5 The disclosed NOV59 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 59C.

Table 59C. BLAST results for NOV59					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 7705375 ref NP_056982.1 (NM_015898)	HIV-1 inducer of short transcripts binding protein [Homo sapiens]	584	548/584 (93%)	548/584 (93%)	e-167
gi 16758916 ref NP_446454.1 (NM_054002)	leukemia/lymphoma related factor [Rattus norvegicus]	569	477/579 (82%)	489/579 (84%)	e-159
gi 6754572 ref NP_034861.1 (NM_010731)	leukemia/lymphoma related factor [Mus musculus]	565	479/579 (82%)	488/579 (83%)	e-125
gi 3599513 gb AAC35368.1 (AF086831)	leukemia/lymohoma related factor cLRF [Gallus gallus]	546	359/578 (62%)	400/578 (69%)	4e-13
gi 2145062 gb AAB58414.1 (AF000561)	TTF-I interacting peptide 21; TIP21; Transcription Termination Factor I Interacting Peptide 21 [Homo sapiens]	590	293/297 (98%)	295/297 (98%)	4e-12

10 Tables 59D-E list the domain descriptions from DOMAIN analysis results against NOV59. This indicates that the NOV59 sequence has properties similar to those of other proteins known to contain this domain.

Table 59D. Domain Analysis of NOV59

gnl|Pfam|pfam00651, BTB, BTB/POZ domain. The BTB (for BR-C, ttk and bab) or POZ (for Pox virus and Zinc finger) domain is present near the N-terminus of a fraction of zinc finger (pfam00096) proteins and in proteins that contain the pfam01344 motif such as Kelch and a family of pox virus proteins. The BTB/POZ domain mediates homomeric dimerisation and in some instances heteromeric dimerisation. The structure of the dimerised PLZF BTB/POZ domain has been solved and consists of a tightly intertwined homodimer. The central scaffolding of the protein is made up of a cluster of alpha-helices flanked by short beta-sheets at both the top and bottom of the molecule. POZ domains from several zinc finger proteins have been shown to mediate transcriptional repression and to interact with components of histone deacetylase co-repressor complexes including N-CoR and SMRT. The POZ or BTB domain is also known as BR-C/Ttk or ZiN

CD-Length = 114 residues, 100.0% aligned

Score = 122 bits (305), Expect = 7e-29

Table 59E. Domain Analysis of NOV59

gnl|Pfam|pfam00096, zf-C2H2, Zinc finger, C2H2 type. The C2H2 zinc finger is the classical zinc finger domain. The two conserved cysteines and histidines co-ordinate a zinc ion. The following pattern describes the zinc finger. #-X-C-X(1-5)-C-X3-#-X5-#-X2-H-X(3-6)-[H/C] Where X can be any amino acid, and numbers in brackets indicate the number of residues. The positions marked # are those that are important for the stable fold of the zinc finger. The final position can be either his or cys. The C2H2 zinc finger is composed of two short beta strands followed by an alpha helix. The amino terminal part of the helix binds the major groove in DNA binding zinc fingers.

CD-Length = 23 residues, 100.0% aligned

Score = 38.5 bits (88), Expect = 0.001

The HIV-1 promoter directs the synthesis of two classes of transcripts, short, non-polyadenylated transcripts and full-length, polyadenylated transcripts. The synthesis of these transcripts is activated by a bipartite DNA element, the inducer of short transcripts or IST, located downstream of the HIV-1 transcriptional start site, while the synthesis of full-length transcripts is activated by the viral activator Tat. Tat binds to the RNA element TAR, which is encoded largely between the two IST half-elements. Upon activation by Tat, the synthesis of short RNAs is repressed. A factor called FBI-1 (for factor that binds to IST) whose binding to wild-type and mutated ISTs correlated well with the abilities of these ISTs to direct the synthesis of short transcripts was identified by Morrison, et al (See Morrison et al., Nucleic Acids Res 1999 Mar 1: 1251-62). FBI-1 contains a POZ domain at its N-terminus and four Kruppel-type zinc fingers at its C-terminus. The C-terminus is sufficient for specific binding, and FBI-1 can form homomers through its POZ domain and, in vivo, through its zinc finger

domain as well. In addition, FBI-1 associates with Tat, suggesting that repression of the short transcripts by Tat may be mediated through interactions between the two factors.

5 The disclosed NOV59 nucleic acid of the invention encoding a HIV-1 inducer of short transcripts binding protein like protein-like protein includes the nucleic acid whose sequence is provided in Table 59A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 59A while still encoding a protein that maintains its HIV-1 inducer of short transcripts binding protein like protein-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences
 10 are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or
 15 derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 19 percent of the bases may be so changed.

20 The disclosed NOV59 protein of the invention includes the HIV-1 inducer of short transcripts binding protein like protein-like protein whose sequence is provided in Table 59B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 59B while still encoding a protein that maintains its HIV-1 inducer of short transcripts binding protein like protein-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up
 25 to about 25 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

30 The above defined information for this invention suggests that this HIV-1 inducer of short transcripts binding protein like protein-like protein (NOV59) may function as a member of a "HIV-1 inducer of short transcripts binding protein like protein family". Therefore, the NOV59 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug

targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

5 The NOV59 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the HIV-1 inducer of short transcripts binding protein like protein-like protein (NOV59) may be useful in gene therapy, and the HIV-1 inducer of short transcripts binding protein like protein-like protein (NOV59) may be useful when administered to a subject in need thereof. By way of nonlimiting
10 example, the compositions of the present invention will have efficacy for treatment of patients suffering from Human Immunodeficiency Virus/ Acquired immune deficiency syndrome, cancer, and inflammatory diseases, or other pathologies or conditions. The NOV59 nucleic acid encoding the HIV-1 inducer of short transcripts binding protein like protein-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications,
15 wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV59 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV59 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-
20 NOVX Antibodies” section below. The disclosed NOV59 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

25 **NOV60**

NOV60 includes three beta tectorin-like proteins disclosed below. The disclosed sequences have been named NOV60a, NOV60b and NOV60c.

NOV60a

A disclosed NOV60a nucleic acid of 1011 nucleotides (also referred to as CG57574-
30 01) encoding a beta tectorin-like protein is shown in Table 60A. The start and stop codons are in bold letters.

Table 60A. NOV60a nucleotide sequence (SEQ ID NO:137).

GATCGAGGCTCAGGCCCTGGAAGGACCGTAAACATTTGGCCAGCTTGGTTTGGATACCTG
GCAGAGACCAGGTTCTGAGAAGCAATGGTGACGAAGGCCCTTTGTCTTGTGGCCATCTTT
GCAGAAGCCTCTGCAAAATCGTGTGCTCCAAATAAAGCAGATGTCATTCTTGTGTTTTGC
TATCCCCAAAACCATCATCACCAAAATCCCCGAGTGTCCCTATGGATGGGAAGTTCATCAG
CTGGCCCTCGGAGGGCTGTGTTACAATGGGGTCCACGAAGGAGTTACTACCAATTTGTG
ATCCCAGATTTATCACCTAAAAACAAGTCTATTGTGGAACCCAGTCTGAGTACAAGCCA
CCTATCTATCACTTCTACAGTCACATCGTTTCCAATGACACCACAGTGATTGTAAAAAAC
CAGCCTGTCAACTACTCCTTCTCCTGCACCTACCACTCCACCTACTTGGTGAACCAGGCT
GCCTTTGACCAGAGTGTCAATTTCTTCCAAAGAATGCCAAGTCTCCATCAAGAAAGAA
GCTCCCTTTGTCTGGAGGCATCCGAAATCGGTTTCAGATCTGTTTGCAGGAGTGAAGCC
AAAGGGTTAAGCATTAGGTTTAAAGTGGTCTTGAACAGCTGTTGGGCCACCCCTCGGCT
GACTTCATGTATCCCTTGCAGTGGCAGCTGATCAACAAGGGCTGCCCCACGGATGAAACC
GTCCCTCGTGCATGAGAATGGGAGAGATCACAGGGCAACCTTCCAATTCAATGCTTTCCGG
TTCCAGAACATCCCCAAACTCTCCAAGTGTGTTACACTGTGAGACGTTTCATCTGCGAC
AGTGAGAAACTCTCCTGCCAGTGACCTGCGATAAACGGAAGCGCTCCTGCGAGACCAG
ACCGGGGGAGTCCCTGGTCTGGAGCTCTCCCTGCGGAATGTTCTCCACCACCTCATCATG
ATGTTGGGGATTGTGCCGTGTTATAGGAGTTAGCCAGGCAGCTGCCGCTCCTCCACCCA
CAATAG

In a search of public sequence databases, the NOV60a nucleic acid sequence, located on chromosome 10 has 428 of 496 bases (86%) identical to a gb:GENBANK-ID:MMBETATEC|acc:X99806.2 mRNA from Mus musculus (Mus musculus mRNA for beta tectorin). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV60a polypeptide (SEQ ID NO:138) encoded by SEQ ID NO:137 has 300 amino acid residues and is presented in Table 60B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV60a has a signal peptide and is likely to be localized plasma membrane with a certainty of 0.6850. The most likely cleavage site for a NOV60a peptide is between amino acids 17 and 18.

Table 60B. Encoded NOV60a protein sequence (SEQ ID NO:138).

MVTKAFVLLAIFAEASAKSCAPNKADVILVFCYPKTIITKIPECPTYGWEVHQALGGLCY
NGVHEGGYYQFVIPDLSPKNKSYCGTQSEYKPPIYHFYSHIVSNDTTVIVKNQPVNYSFS
CTYHSTYLVNQAAFDQSVNFLPKNAKFSIKKEAPFVLEASEIGSDLFAGVEAKGLSIRFK
VVLNSCWATPSADFMYPQLQWQLINKGCPTDETVLVHENGDRHRATFQFNAFRFQNPILS
KVWLHCETFICDSEKLSCPVTCDKRKRLRLDQTGGVLVVELSLRNVLHHLIMMLGICAVL

A search of sequence databases reveals that the NOV60a amino acid sequence has 149 of 174 amino acid residues (85%) identical to, and 155 of 174 amino acid residues (89%) similar to, the 329 amino acid residue ptrn:SPTREMBL-ACC:O08524 protein from Mus musculus (Mouse) (BETA TECTORIN). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV60a is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:MMBETATEC|acc:X99806.2) a closely

related *Mus musculus* mRNA for beta tectorin homolog in species *Mus musculus* :cochleae. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

5 NOV60b

A disclosed NOV60b nucleic acid of 1012 nucleotides (also referred to as CG57574-02) encoding a beta tectorin-like protein is shown in Table 60C. The start and stop codons are in bold letters.

Table 60C. NOV60b nucleotide sequence (SEQ ID NO:139).

```
AGAGACCAGGTTCTGAGAAGCAATGGTGACGAAGGCCTTTGTCTTGTGGCCATCTTTGCAGAAGCCTCT
GCAAAATCGTGTGCTCCAAATAAAGCAGATGTCAATCTTGTGTTTGTCTATCCCAAAACCATCATACCA
AAATCCCCGAGTGTCCCTATGGATGGGAAGTTCATCAGCTGGCCCTCGGAGGGCTGTGTTACAATGGGGT
CCACGAAGGAGGTACTACCAATTGTGTATCCAGATTTATCACCTAAAAACAAGTCTATTGTGGAACC
CAGTCTGAGTACAAGCCACCTATCTATCACTTCTACAGTCACATCGTTTCCAATGACGCCACAGTGATTG
TAAAAAACCAGCCTGTCAACTACTCCTTCTCCTGCACCTACCCTCCACCTACTTGGTGAACCAGGCTGC
CTTTGACCAGAGAGTGGCCACTGTTACGTGAAGAACGGGAGCATGGGCACATTGAGAGCCAACTGTCT
CTCAACTTCTACACTAATGCCAAGTCTCCATCAAGAAAGAAGCTCCCTTTGTCTGGAGGCATCCGAAA
TCGGTTCAGATCTGTTTGAGGAGTGAAGCCAAAGGGTTAAGCATTAGGTTTAAAGTGGTCTTGAACAG
CTGTTGGGGCCACCCCTCGGCTGACTTCATGTATCCCTTGAGTGGCAGCTGATCAACAAGGGCTGCCCC
ACGGATGAAACCGTCCCTCGTGCATGAGAATGGGAGAGATCACAGGGCAACCTTCCAATTCAATGCTTTCC
GGTTCAGAACATCCCCAACTCTCCAAGGTGTGGTTACACTGTGAGACGTTTCTGCGACAGTGAGAA
ACTCTCTGCCCCAGTGACCTGCGATAAACCGAAGCGCCTCCTGCGAGACCAGACCGGGGGAGTCTGTGTC
GTGGAGCTCTCCCTGCGGAGCAGGGGATTTTCCAGTCTCTATAGCTTCTCAGATGTTCTCCACCACCTCA
TCATGATGTTGGGGATTGTGCCGTGTTATAG
```

In a search of public sequence databases, the NOV60b nucleic acid sequence, located on chromosome 10 has 887 of 1012 bases (87%) identical to a gb:GENBANK-ID:MMBETATEC|acc:X99806.2 mRNA from *Mus musculus* (*Mus musculus* mRNA for beta tectorin). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV60b polypeptide (SEQ ID NO:140) encoded by SEQ ID NO:139 has 329 amino acid residues and is presented in Table 60D using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV60b has a signal peptide and is likely to be localized extracellularly with a certainty of 0.6850. The most likely cleavage site for a NOV60b peptide is between amino acids 17 and 18.

Table 60D. Encoded NOV60b protein sequence (SEQ ID NO:140).

```
MVTKAFVLLAIFAEASAKSCAPNKADVILVFCYPKTIITKIPECPYGWEVHQLALGGLCY
NGVHEGGYYQFVIPDLSPKNKSYCGTQSEYKPIYHFYSHIVSNDATVIVKNQPVNYSFS
CTYHSTYLVNQAAFDQRVATVHVKNKSMGTTFESQLSLNFYTNAKFSIKKEAPFVLEASEI
GSDLFAGVEAKGLSIRFKVVLNSCWATPSADFMYPQLQWQLINKGCPTDETIVLVHENGDRH
RATFQFNAFRFQNI PKLSKVWLHCETFICDSEKLSCPVTCDKRKRLLRDQTGGVLVVELS
LRSRGFSSLYSFSVDLHHLIMMLGICAVL
```

A search of sequence databases reveals that the NOV60b amino acid sequence has have 310 of 329 amino acid residues (94%) identical to, and 317 of 329 amino acid residues (96%) similar to, the 329 amino acid residue ptnr:SPTREMBL-ACC:O08524 protein from Mus musculus (Mouse) (BETA TECTORIN). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV60b is predicted to be expressed in at least the ear. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

NOV60c

A disclosed NOV60c nucleic acid of 1012 nucleotides (also referred to as CG57574-02) encoding a beta tectorin-like protein is shown in Table 60E. The start and stop codons are in bold letters.

Table 60E. NOV60c nucleotide sequence (SEQ ID NO:141).
AGAGACCAGGTTCTGAGAAGCA ATGGT GACGAAGGCC TTTGTCTTGT TGGCCATC TTTGC AGAAGCCTCTGCAAAATCGTGTGCTCCAATAAAGCAGATGTCAT CTTGTGTTT GTCTA TCCCAAACCATCATCACCAAATCCCCGAGTGTCCCTATGGATGGGAAGTTCATCAGCT GGCCCTCGGAGGGCTGTGTTACAATGGGGTCCACGAAGGAGGT TACTACCAATTT GTGAT CCCAGATTTATCACCTAAAAACAAGTCTTATGTGGAACCCAGTCTGAGTACAAGCCACC TATCTATCACTTCTACAGTCACATCGTTTCCAATGACACCACAGTGATTGTA AAAAACCA GCCTGTCAACTACTCCTTCTCCTGCACCTACCACTCCACCTACTTGGTGAACCAGGCTGC CTTTGACCAGAGAGTGGCCACTGTTACAGTGAAGAACGGGAGCATGGGCACATTTGAGAG CCAAC TGTCTCTCAACTTCTACACTAATGCCAAGTTC CCATCAAGAAAGAAAGCTCCCTT TGTCTGGAGGCATCGGAAATCGGTT CAGATCTGTTT GCAGGAGTGGAAAGCCAAAGGGTT AAGCATTAGGTTTAAAGTGGTCTTGAACAGCTGTTGGGCCACCCCTCGGCTGACTTCAT GTATCCCTTGCAGTGGCAGCTGATCAACAAGGGCTGCCCCACGGATGAAACCGTCCCTCGT GCATGAGAATGGGAGAGATCACAGGGCAACCTTCCAATTCAATGCTTCCGGTTCCAGAA CATCCCCAACTCTCCAAGGTGTGGTTACACTGTGAGACGTT CATCTGCGACAGTGAGAA ACTCTCCTGCCAGTGACCTGCGATAAACGGAAGCGCCTCCTGCGAGACCAGACCGGGG AGTCCTGGTTCGTGGAGCTCTCCCTGCGGAGCAGGGGATTTCCAGTCTCTATAGCTTCTC AGATGTTCTCCACCACCTCATCATGATGTTGGGGATTTGTGCCGTGTTAT AG

In a search of public sequence databases, the NOV60c nucleic acid sequence, located on chromosome 10 has 887 of 1012 bases (87%) identical to a gb:GENBANK-ID:MMBETATEC|acc:X99806.2 mRNA from Mus musculus (Mus musculus mRNA for beta tectorin). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV60c polypeptide (SEQ ID NO:142) encoded by SEQ ID NO:141 has 329 amino acid residues and is presented in Table 60F using the one-letter amino acid

code. Signal P, Psort and/or Hydropathy results predict that NOV60c has a signal peptide and is likely to be localized plasma membrane with a certainty of 0.6850. The most likely cleavage site for a NOV60c peptide is between amino acids 17 and 18.

Table 60F. Encoded NOV60c protein sequence (SEQ ID NO:142).

MVTKAFVLLAIFAEASAKSCAPNKADVILVFCYPKTIITKIPECYPGWVHQLALGGLCY
NGVHEGGYYQFVIPDLSPKNKSYCGTQSEYKPPIYHFYSHIVSNDTTVIVKNQPVNYSFS
CTYHSTYLVNQAAFDQRVATVHVKNKSGMTFESQLSLNFYTNAKFSIKKEAPFVLEASEI
GSDLFAGVEAKGLSIRFKVVLNSCWATPSADFMYPQLQWLINKGCPTDETVLVHENGDRH
RATFQFNAFRFQNIPLKLSKVWLHCETFCIDSEKLSCPVTCDKRKRLLRDQTGGVLVVELS
LRSRGFSSLYSFSVDVLHHLIMMLGICAVL

5 A search of sequence databases reveals that the NOV60c amino acid sequence has 310 of 329 amino acid residues (94%) identical to, and 317 of 329 amino acid residues (96%) similar to, the 329 amino acid residue ptnr:SPTREMBL-ACC:O08524 protein from Mus musculus (Mouse) (BETA TECTORIN)Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

10 NOV60c is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:MMBETATEC|acc:X99806.2) a closely related Mus musculus mRNA for beta tectorin homolog in species Mus musculus :cochleae. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources,
15 Literature sources, and/or RACE sources.

The disclosed NOV60c polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 60G.

Table 60G. BLAST results for NOV60c

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17158035 ref NP_478129.1 (NM_058222)	tectorin beta [Homo sapiens]	329	295/330 (89%)	297/330 (89%)	e-167
gi 7363457 ref NP_033374.2 (NM_009348)	tectorin beta; [b]-tectorin [Mus musculus]	329	280/330 (84%)	289/330 (86%)	e-159
gi 1729889 sp P54097 TECB CHICK	BETA-TECTORIN PRECURSOR	329	220/330 (66%)	260/330 (78%)	e-125
gi 13385494 ref NP_080265.1 (NM_025989)	RIKEN cDNA 2310037I18 [Mus musculus]	531	63/208 (30%)	99/208 (47%)	4e-13

gi 12844889 dbj BAB 26538.1 (AK009843)	homolog to PANCREATIC SECRETORY GRANULE MEMBRANE MAJOR GLYCOPROTEIN GP2 PRECURSOR (PANCREATIC ZYMOGEN GRANULE MEMBRANE PROTEIN GP-2) (ZAP75)~putative [Mus musculus]	573	62/200 (31%)	96/200 (48%)	4e-12
--	---	-----	-----------------	-----------------	-------

Table 60H lists the domain descriptions from DOMAIN analysis results against NOV60c. This indicates that the NOV60c sequence has properties similar to those of other proteins known to contain this domain.

5

<p>Table 60H. Domain Analysis of NOV60c</p> <p>gnl Smart smart00241, ZP, Zona pellucida (ZP) domain; ZP proteins are responsible for sperm-adhesion fo the zona pellucida. ZP domains are also present in multidomain transmembrane proteins such as glycoprotein GP2, uromodulin and TGF-beta receptor type III (betaglycan).</p> <p>CD-Length = 253 residues, 98.0% aligned Score = 119 bits (297), Expect = 3e-28</p>

Legan et al. (1997) cloned mouse alpha- and beta-tectorins. The mouse beta-tectorin gene encodes a 320-amino acid protein containing a hydrophobic secretory signal sequence and 4 potential N-glycosylation sites. Both alpha- and beta-tectorin contain a zona pellucida domain, but otherwise are not homologous.

10

To identify genes expressed in the vertebrate inner ear, Heller et al. (1998) established an assay that allowed rapid analysis of the differential expression pattern of mRNAs derived from an auditory epithelium-specific cDNA library. They performed subtractive hybridization to create an enriched probe, which was then used to screen the cDNA library. After digoxigenin-labeled antisense cRNAs had been transcribed from hybridization-positive clones, they conducted in situ hybridization on slides bearing cryosections of late embryonic chicken heads, bodies, and cochleas. They found 12 proteins whose mRNAs were specifically or highly expressed in the chicken's inner ear; the remainder encoded proteins that occur more widely. They identified proteins that had previously been described as expressed in the inner ear, such as beta-tectorin, calbindin (CALB1; 114050), and type II collagen (COL2A1; 120140). A second group of proteins abundant in the inner ear included 5 additional types of

15

20

collagen. A third group, including COCH5B2 (COCH; 603196) and ear-specific connexin, comprised the proteins whose human equivalents are candidates to account for hearing disorders. This last group also included proteins expressed in 2 cells types unique to the inner ear, homogene cells and cells of the tegmentum vasculosum.

5 The disclosed NOV60c nucleic acid of the invention encoding a beta tectorin-like protein includes the nucleic acid whose sequence is provided in Table 60A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 60A while still encoding a protein that maintains its beta tectorin-like activities and physiological functions, or a fragment of
10 such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example,
15 modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 13 percent of the bases may be so changed.

20 The disclosed NOV60 protein of the invention includes the beta tectorin-like protein whose sequence is provided in Table 60B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 60B while still encoding a protein that maintains its beta tectorin-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up
25 to about 6 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this beta tectorin-like protein (NOV60) may function as a member of a “beta tectorin family”. Therefore, the
30 NOV60 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene

delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV60 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the beta tectorin-like protein (NOV60) may be useful in gene therapy, and the beta tectorin-like protein (NOV60) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from hearing loss, or other pathologies or conditions. The NOV60 nucleic acid encoding the replacement-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV60 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV60 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV60 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV61

A disclosed NOV61 nucleic acid of 3802 nucleotides (also referred to as CG57505-01) encoding a KIAA1125-like protein is shown in Table 61A. The start and stop codons are in bold letters.

Table 61A. NOV61 nucleotide sequence (SEQ ID NO:143).

GGGCAGCCAATCGGGGATGAGCTTTTATTAGGCGGCCAGATCATCAGCCGAAGTGCCAAA
CCCTTTTCTGTGAGAACTAGGAGCCTGTCCCTCCATGTTTTATAAGTATTGACATTACAC
AGTGTTAACA**ATGC**ATCCACAGAGCTTGGCTGAAGAGGAAAATAAAACAGAACAGGAGGT
GGTAGAGGGCATGGATATCTCTACTCGCTCCAAAGATCCTGGCTCTGCAGAGAGAACAGC
CCAGAAAAGAAAGTTCCCCAGCCCTCCACATTTCTTCCAATGGCCACTCGCCGCAGGACAC
ATCAACAAGCCCCATTAAAAAGAAAAAGAAACCTGGCTTACTGAACAGTAACAATAAGGA
GCAGTCAGAACTAAGACATGGTCCGTTTTACTATATGAAGCAGCCACTCACCACAGACCC
TGTTGATGTTGTACCGCAGGATGGACGGAATGATTTCTACTGCTGGGTTTGTACCCGGA
AGGCCAAGTCCTTTGCTGTGAGCTCTGTCCCGGGTTTATCACGCTAAGTGTCTGAGACT
GACATCGGAACCAGAGGGGGACTGGTTTTGTCTTGAATGTGAGAAAATTACAGTAGCAGA
ATGCATCGAGACCCAGAGTAAAGCCATGACAATGCTCACCATTGAACAGTTATCTACCT

GCTCAAGTTTGCCATTTCAGAAAATGAAACAGCCAGGGACAGATGCATTCCAGAAGCCCCGT
TCCATTGGAACAGCACCCTGACTATGCGGAATACATCTTCCATCCAATGGACCTTTGTAC
ATTGGAAAAGAATGCGAAAAAGAAAATGTATGGCTGCACAGAAGCCTTCCTGGCTGATGC
AAAGTGGATTTTGCACAACCTGCATCATTTATAATGGGGGAAATCACAAATTGACGCAAAAT
AGCGAAAGTAGTCATCAAAATCTGTGAACATGAGATGAATGAAATCGAAGTATGTCCAGA
ATGTTATCTAGCTGCTTGCCAAAAACGAGATAACTGGTTTTGTGAGCCTTGTAGCAATCC
ACATCCTTTGGTCTGGGCCAACTGAAGGGGTTTCCATTCTGGCCTGCAAAAGCTCTAAG
GGATAAAGACGGGCAGGTTCGATGCCCCGATTCTTTGGACAACATGACAGGGCCTGGGTTCC
AATAATAATTGCTACCTCATGTCTAAAGAAATTCCTTTTTCTGTGAAAAAGACTAAGAG
CATCTTCAACAGTGCCATGCAAGAGATGGAGGTTTACGTGGAGAACATCCGCAGGAAGTT
TGGGGTTTTTAATTACTCTCCATTTAGGACACCCTACACACCCAACAGCCAGTATCAAAAT
GCTGCTCGATCCCAACCCACGCGCCGGCACTGCCAAGATAGACAAGCAGGAGAAGGT
CAAGCTCAACTTTGACATGACGGCATCCCCAAGATCCTGATGAGCAAGCCTGTGCTGAG
TGGGGGCACAGGCCCGCGGATTTCTTGTGGATATGCCGCGCTCCCCATGAGCACAAA
CTCTTCTGTGCACACGGGCTCCGACGTGGAGCAGGATGCTGAGAAGAAGGCCACGTGAG
CCACTTCAGTGCAGCGAGGAGTCCATGGACTTCTGGATAAGAGCACAGCTTACCAGC
CTCCACCAAGACGGGACAAGCAGGGAGTTTATCCGGCAGCCCAAAGCCCTTCTCTCTCA
ACTGTCAGCTCCTATCACGACGAAAACGGACAAAACCTCCACCACCGGCAGCATCTTGAA
TCTTAACCTGGATCGAAGCAAAGCTGAGATGGATTTGAAGGAGCTGAGCGAGTCCGTCCA
GCAACAGTCCACCCCTGTTCTCTCATCTCTCCCAAGCGCCAGATTTCGTAGCAGGTTCCA
GCTGAATCTTGACAAGACCATAGAGAGTTGCAAAGCACAAATAGGCATAAATGAAATCTC
GGAAGATGTCTATACGGCCGTAGAGCACAGCGATTCCGGAGGATTCGAGAAGTCAGATAG
TAGCGATAGTGAGTATATCAGTGATGATGAGCAGAAGTCTAAGAACGAGCCAGAAGACAC
AGAGACATAAGAAGGTTGTGATGGACAAAGAGCCATCTGCTGTTAAAAAAGGCCCAA
GCCTACAAACCCAGTGGAGATTAAAGAGGAGCTGAAAAGCACGTACCAGCCAGCGAGAA
GGCAGACCCTGGAGCAGTCAAGGACAAGGCCAGCCCTGAGCCTGAGAAGGACTTTTCCGA
AAAGGCAAAACCTTACCTCACCCATAAAGGATAAACTGAAGGGAAAAGATGAGACGGA
TTCCCCAACAGTCCATTTGGGCCTGGACTCTGATTTCAGAGAGCGAACTTGTATAGATTT
AGGAGAAGACCATTCTGGGCGGGAGGGTCGAAAAATAAGAAGGAACCCAAAGAACCATC
TCCCCAACAGGATGTTGTAGGTAATACTCCACCATCCACGACGGTGGGCAGCCATTCTCC
CCCCGAAACACCGGTGCTCACCGCTCTTCCGCCCAAACCTCCGCGGCTGGCGCCACAGC
CACCACAGCACGTCTCTCCAGGTCACCGTCACGGCCCCGGCCCCCGCCGACAGGAAG
CCCAGTGAAAAAGCAGAGGCCGCTTTTACCGAAGGAGACTGCCCGGCCGTGACGGGGT
CGTGTGGAACCTCATCAAGTAAGTTTCAAACGTCTCTCCAAAAGTGGCACATGCAGAAGAT
GCAGCGTCAGCAGCAGCAGCAGCAGCAGCAAAACCAGCAGCAGCAGCCTCAGTCTTCCCA
GGGGACGAGATATCAGACCAGACAGGCTGTGAAAGCTGTCCAGCAGAAGGAGATCACACA
GAGCCCATCCACGTCCACCATCACCTGGTGACCAGCACACAGTCATCGCCCCCTGGTCAC
GAGCTCGGGGTCCATGAGCACCTTGTGTCTCAGTCAACGCTGACCTGCCCATCGCCAC
TGCCCTCAGCTGATGTCGCCGCTGATATTGCCAAGTACACTAGCAAAATGATGGATGCAAT
AAAAGGAACAATGACAGAAATATACAACGATCTTTCTAAAAACACTACTGGAAGCACAAT
AGCTGAGATTTCGAGGCTGAGGATCGAGATAGAGAAGCTCCAGTGGCTGCACCAGCAAGA
GCTCTCCGAAATGAAACACAACCTTAGAGCTGACCATGGCGGAGATGCGGCAGAGCCTGGA
GCAGGAGCGGGACCGGCTCATCGCCGAGGTGAAGAAGCAGCTGGAGTTGGAGAAGCAGCA
GGCGGTGGATGAGACCAAGAAGAAGCAGTGGTGCGCCAACTGCAAGAAGGAGGCCATCTT
TTACTGCTGTTGGAACACTAGCTACTGTGACTACCCCTGCCAGCAAGCCCACTGGCCTGA
GCACATGAAGTCTGCACCCAGTCAGTACTGCTCCTCAGCAGGAAGCGGATGCTGAGGT
GAACACAGAAAACATAAATAAGTCTCCAGGGGAGCTCCTCGAGCACACAATCAGCACC
TTCAGAAACGGCCAGCGCCTCCAAAGAGAAGGAGACGTGAGCTGAGAAAAGCAAGGAGAG
TGGCTCGACCCCTTGACCTTTCTGGCTCCAGAGAGACGCCCTCCTCCATTCTCTTAGGCTC
CAACCAAGGCTCTGACCATTCCCGGAGTAATAAATCCAGTTGGAGCAGCAGTGTGAGAA
GAGGGGATCGACACGTTCCGATCACAACACCAGTACCAGCACGAAGAGCCTCTCCCGAA
AGAGTCTCGGCTGGACACCTTCTGGGACTAGCAGTGAATCGGGACACAAACCACCCACCC
CATTGGGAGAAAAACCCAGACGCCAGGAAAAAGAAGAAACAACAAAGGCAGGAGAACAGCC
ACTTTCAGACTTGAAAATGACAAAACCTCAGTTGAGCCTGAGCCCCGGCGCGGGGCT
GCTACACTA

In a search of public sequence databases, the NOV61 nucleic acid sequence, located on chromosome 20 has 3797 of 3827 bases (99%) identical to a gb:GENBANK-
ID:AB032951|acc:AB032951.1 mRNA from Homo sapiens (Homo sapiens mRNA for

KIAA1125 protein, partial cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV61 polypeptide (SEQ ID NO:144) encoded by SEQ ID NO:143 has 1206 amino acid residues and is presented in Table 61B using the one-letter amino acid code.

5 Signal P, Psort and/or Hydropathy results predict that NOV61 has no signal peptide and is likely to be localized in the nucleus with a certainty of 0.7000.

Table 61B. Encoded NOV61 protein sequence (SEQ ID NO:144).	
MHPQSLAEEEEIKTEQEVVEGMDISTRKDPGSAERTAQKRKFSPPHSSNGHSPQDTSTS	
PIKKKKKPGLLNSNNKEQSELRHGPFYYMKQPLTTDPVDVVPQDGRNDFYCWCVCHREGQV	
LCCELCPRVYHAKCLRLTSEPEGDWFCPECEKITVAECIETQSKAMTMLTIEQLSYLLKF	
AIQKMKQPGTDAFQKVPLEQHPDYAEYIFHPMDLCTLEKNAKKKMYGCTEAFADAKWI	
LHNCIIYNGGNHKLTIQIAKVVIKICEHEMNEIEVCPECYLAACQKRDNWFCEPCSNPHPL	
VWAKLKGFPFWPAKALRDKDGQVDARFFGQHDRAWVPINNCYLMSKEIPFSVKKTKSIFN	
SAMQEMEYVENIRRKFGVFNYSPPFTPTPNQYQMLLDPTNPSAGTAKIDKQEKVKLN	
FDMTASPKILMSKPVLSGGTGRRISLSDMPRSPMSTNSSVHTGSDVEQDAEKKATSSHFS	
ASEESMDFLDKSTASPASTKTGQAGSLSGSPKPFSPQLSAPITTKTDKTSTTGSILNLL	
DRSKAEMDLKELSESQVQQSTPVPPLISPKRQIRSRFQLNLDKTIESCQAQLGINEISEDV	
YTAVEHSDSESEKSDSSSEYISDDEQKSKNEPEDTEDKEGCQMDKEPSAVKKPKPTN	
PVEIKEELKSTSPASEKADPGAVKDKASPEPEKDFSEKAKPSHPKIDKLGKDETDSP	
VHLGLDSDSESELVIDLGEDHSGREGKKNKKEPKPSPKQDVVGKTPPSTTVGSHSPPET	
PVLTRSSAQTSAGATATTSTSTSTVTVTAPAPAATGSPVKKQRPLLPKETAPAVQRVWN	
SSSKFQTSSQKWHMQMKMRQQQQQQQQNQQQQPQSSQGTQYQTRQAVKAVQKEITQSPS	
TSTITLVTSTQSSPLVTSSGSMSTLVSSVNADLPATASADVAADIAKYTSKMMDAIGT	
MTEIYNDLSKNTTGSTIAEIRRLRIEIEKLQWLHQQELSEMKNLELTMAEMRQSLER	
DRLIAEVKKQLELEKQQAQVDETCKKQWCANCKKEAIFYCCWNTSYCDYPCQQAHWPEHMK	
SCTQSATAPQQEADAENVNTETLNKSSQSSSTQSAPSETASASKEKETSAEKSKESGST	
LDLGSRETPSSILLGSNQSDHSRSNKSSWSSSDEKRGSTRSDHNTSTSTKSLLPKESR	
LDTFWD	

A search of sequence databases reveals that the NOV61 amino acid sequence has 1201
 10 of 1201 amino acid residues (100%) identical to, and 1201 of 1201 amino acid residues (100%) similar to, the 1205 amino acid residue ptrn:SPTREMBL-ACC:Q9ULU4 protein from Homo sapiens (Human) (KIAA1125 PROTEIN). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV61 is expressed in at least Brain, Lymphoid tissue, Kidney, Whole Organism,
 15 Bone Marrow, Prostate, Lung, Lung Pleura, Retina. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV61 polypeptide has homology to the amino acid sequences shown
 20 in the BLASTP data listed in Table 61C.

Table 61C. BLAST results for NOV61					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 6329749 dbj BAA86439.1 (AB032951)	KIAA1125 protein [Homo sapiens]	1205	1201/1201 (100%)	1201/1201 (100%)	0.0
gi 17980969 gb AAL50790.1 AF454056.1 (AF454056)	se14-3r protein [Homo sapiens]	995	992/1041 (95%)	994/1041 (95%)	0.0
gi 14786224 ref XP_012932.3 (XM_012932)	similar to protein kinase C binding protein 1 (H. sapiens) [Homo sapiens]	949	932/933 (99%)	933/933 (99%)	0.0
gi 11385648 gb AAG34905.1 AF273045.1 (AF273045)	CTCL tumor antigen se14-3 [Homo sapiens]	764	762/810 (94%)	763/810 (94%)	0.0
gi 13677199 emb CAC19781.1 (AL031666)	dJ569M23.1. 1 (protein kinase C binding protein 1, isoform 1 (KIAA1125)) [Homo sapiens]	521	520/520 (100%)	520/520 (100%)	0.0

Table 61D-F lists the domain descriptions from DOMAIN analysis results against NOV61. This indicates that the NOV61 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 61D. Domain Analysis of NOV61

[gnl|Smart|smart00297](#), BROMO, bromo domain
CD-Length = 109 residues, 94.5% aligned
Score = 81.3 bits (199), Expect = 3e-16

Table 61E. Domain Analysis of NOV61

[gnl|Pfam|pfam00628](#), PHD, PHD-finger. PHD folds into an interleaved type of Zn-finger chelating 2 Zn ions in a similar manner to that of the RING and FYVE domains.
CD-Length = 49 residues, 93.9% aligned
Score = 56.6 bits (135), Expect = 8e-09

Table 61F. Domain Analysis of NOV61

gnl|Pfam|pfam01753, zf-MYND, MYND finger.

CD-Length = 38 residues, 100.0% aligned

Score = 45.1 bits (105), Expect = 2e-05

Mukai and Ono (1994) isolated a cDNA for a protein kinase, designated PKN by them, from a human hippocampus cDNA library. The putative 942-amino acid protein has leucine zipper-like sequences at its amino terminus and contains a domain with strong similarity to that of the protein kinase C family. Ubiquitous expression in human tissues was shown.

5 Antisera detected a 120-kD recombinantly expressed protein on Western blots. The protein showed intrinsic protein kinase activity that was abolished by a mutation in the predicted ATP binding site.

Palmer et al. (1994) used degenerate PCR to isolate 3 novel members of the closely related protein kinase C (PKC) family, termed PRK1, PRK2 (602549), and PRK3. Palmer et al. (1995) cloned a full-length cDNA of PRK1 from a human fetal brain library. Using Northern blot and RT-PCR analyses Palmer et al. (1995) detected expression of PRK1 in all tissues and cell lines tested.

15 In a study of proteins that bind to the rho GTPase (see Ridley and Hall, 1992), Amano et al. (1996) discovered 1 protein that had partial amino acid sequences identical to PKN. They found that rho binds directly to a polybasic region of the N-terminal regulatory domain that precedes the leucine zipper-like motif. The authors speculated that through this activity, PKN may mediate the rho-dependent signaling pathway.

Bartsch et al. (1998) used fluorescence in situ hybridization to map the PRKCL1 gene to 19p13.1-p12 and radiation hybrid mapping to localize the gene in subband 19p12. By segregation analysis, they mapped the corresponding mouse gene (Prkcl1) to chromosome 8.

The disclosed NOV nucleic acid of the invention encoding a KIAA1125-like protein includes the nucleic acid whose sequence is provided in Table 61A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 61A while still encoding a protein that maintains its KIAA1125-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include

chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 1 percent of the bases may be so changed.

The disclosed NOV61 protein of the invention includes the KIAA1125-like protein whose sequence is provided in Table 61B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 61B while still encoding a protein that maintains its KIAA1125-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 0 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this KIAA1125-like protein (NOV61) may function as a member of a “KIAA1125 family”. Therefore, the NOV61 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV61 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the KIAA1125-like protein (NOV61) may be useful in gene therapy, and the KIAA1125-like protein (NOV61) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from Cardiomyopathy, Atherosclerosis, Hypertension, Congenital heart defects, Aortic stenosis, Atrial septal defect (ASD), Atrioventricular (A-V) canal defect, Ductus arteriosus, Pulmonary stenosis, Subaortic stenosis, Ventricular septal defect (VSD), valve diseases, Tuberous sclerosis, Scleroderma, Obesity, Transplantation, Aneurysm, Fibromuscular dysplasia, Stroke, Anemia, Bleeding disorders, Adrenoleukodystrophy, Congenital Adrenal

Hyperplasia, Diabetes, Von Hippel-Lindau (VHL) syndrome, Pancreatitis, Hyperparathyroidism, Hypoparathyroidism, SIDS, Endometriosis, Fertility, Xerostomia, Hypercalcaemia, Ulcers, Cirrhosis, Inflammatory bowel disease, Diverticular disease, Hirschsprung's disease, Crohn's Disease, Appendicitis, Hemophilia, hypercoagulation, Idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, Graft versus host, Ataxia-telangiectasia, Hemophilia, Lymphedema, Tonsillitis, Osteoporosis, Arthritis, Ankylosing spondylitis, Scoliosis, Tendinitis, Muscular dystrophy, Lesch-Nyhan syndrome, Myasthenia gravis, Dental disease and infection, Alzheimer's disease, Tuberous sclerosis, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, Growth and reproductive disorders, Endocrine dysfunctions, Systemic lupus erythematosus, Asthma, Emphysema, ARDS, Pharyngitis, Laryngitis, Hearing loss, Tinnitus, Psoriasis, Actinic keratosis, Tuberous sclerosis, Acne, Hair growth, alopecia, pigmentation disorders, cystitis, incontinence, Renal artery stenosis, Interstitial nephritis, Glomerulonephritis, Polycystic kidney disease, Systemic lupus erythematosus, Renal tubular acidosis, IgA nephropathy, Vesicoureteral reflux, glaucoma, blindness, and Hypothyroidism, or other pathologies or conditions. The NOV61 nucleic acid encoding the KIAA1125-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV61 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV61 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV61 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV62

NOV62 includes two zinc finger BOP-like proteins disclosed below. The disclosed sequences have been named NOV62a and NOV62b.

NOV62a

A disclosed NOV62a nucleic acid of 1629 nucleotides (also referred to as CG57473-01) encoding a zinc finger BOP-like protein is shown in Table 62A. The start and stop codons are in bold letters.

5

Table 62A. NOV62a nucleotide sequence (SEQ ID NO:145).
<p>TATAGTCTTGCTCTTTGGGATGCTGAAGGTGCTGAAATAGCAATGACAAGAGACTTGGCT CAGTGTTAAATAACTGCCGCGCTGGCCTGACAGTCTCTGAGATGACAATAGGGAGAATGG AGAACGTGGAGGTCTTCACCGCTGAGGGCAAAGGAAGGGGTCTGAAGGCCACCAAGGAGT TCTGGGCTGCAGATATCATCTTTGCTGAGCGGGCTTATTCGCGAGTGGTTTTTGACAGCC TTGTTAATTTTGTGTGCCACACCTGCTTCAAGAGGCAGGAGAAGCTCCATCGCTGTGGGC AGTGCAAGTTTGGCCATTACTGCGACCGCACCTGCCAGAAGGATGCTTGGCTGAACCACA AGAATGAATGTTTCGGCCATCAAGAGATATGGGAAGGTGCCCAATGAGAACATCAGGCTGG CGGCGCGCATCATGTGGCGGGTGGAGAGAGAAGGCACCGGGCTCACGGAGGGCTGCCTGG TGTCCGTGGACGACTTGCAGAACCACGTGGAGCACTTTGGGGAGGAGGAGCAGAAGGACC TGCGGGTGGACGTGGACACATTCTTGCACTTGGCCGCGCAGAGCCAGCCGTTTCAGCA TGCAGTACATCTCGCACATCTTCGGAGTGATTAACCTGCAACGGTTTTACTCTCAGTGATC AGAGAGCCCTGCAGGCCGTGGGCGTAGGCATCTTCCCAACCTGGGCCCTGGTGAACCATG ACTGTTGGCCCAACTGTACTGTCTATTTTAAACAATGGCAATCATGAGGCAGTGAATCCA TGTTTCATACCCAGATGAGAATTGAACCTGCGGGCCCTAGGCAAGATCTCAGAAGGAGAGG AGCTGACTGTGTCTTATATCGACTTCTCAACGTTAGTGAAGAACGCAAGAGGCAGCTGA AGAAGCAGTACTACTTTGACTGCACATGTGAACACTGCCAGAAAAAAGTGAAGGATGACC TCTTCTCGGGGTGAAAGACAACCCCAAGCCCTCTCAGGAAGTGGTGAAGGAGATGATAC AATTCTCCAAGGATACATTGGAAAAGATAGACAAGGCTCGTTCCGAGGGTTTGTATCATG AGGTTGTGAAATTATGCCGGGAGTGCTTGGAGAAGCAGGAGCCAGTGTGCTGACACCA ACATGTACATGCTGCGGATGCTGAGCATTGTTTCGGAGGTCCTTTCTACCTCCAGGCCT TTGAGGAGGCCTCGTTCTATGCCAGGAGGATGGTGGACGGCTATATGAAGCTCTACCACC CCAACAATGCCCAACTGGGCATGGCCGTGATGCGGGCAGGGCTGACCAACTGGCACGCTG GTAACATTGAGGTGGGGCACGGGATGATCTGCAAGCCATGCCATTCTCTGGTGACAC ACGGACCCTCCACCCCATCACTAAGGACTTAGAGGCCATGCGGGTGCAGACGGAGATGG AGCTACGCATGTTCCGCCAGAACGAATTCATGTACTACAAGATGCGCGAGGCTGCCCTGA ACAACCAGCCCATGCAGGTCATGGCCGAGCCAGCAATGAGCCATCCCAGCTCTGTTCC ACAAGAAGCAATGAGGACTGCCAGTGGAGGAGGGGCGATGTGGCTGGGGAGCTAGGGAG AGACTCTGG</p>

In a search of public sequence databases, the NOV62a nucleic acid sequence, located on chromosome 2 has 1392 of 1573 bases (88%) identical to a gb:GENBANK-
ID:MMU76373|acc:U76373.2 mRNA from Mus musculus (Mus musculus skm-BOP1 (Bop)
10 mRNA, complete cds). Public nucleotide databases include all GenBank databases and the
GeneSeq patent database.

The disclosed NOV62a polypeptide (SEQ ID NO:146) encoded by SEQ ID NO:145
has 490 amino acid residues and is presented in Table 62B using the one-letter amino acid
code. Signal P, Psort and/or Hydropathy results predict that NOV62a has no signal peptide
15 and is likely to be localized in the cytoplasm with a certainty of 0.6500.

Table 62B. Encoded NOV62a protein sequence (SEQ ID NO:146).
<p>MTIGRMENVEVFTAEGKGRGLKATKEFWAADIIFAERAYSAVVFDSLVNFVCHTCFKRQE KLHRCGQCKFAHYCDRTCQKDAWLNHKNECSAISKRYGKVPNENIRLAARIMWRVEREGTG</p>

LTEGCLVSVDDLQNHVEHFGEEEQKDLRVDVDTFLQYWPPQSQPFSMQYISHIFGVINCN
GFTLSDQRGLQAVGVGIFPNLGLVNHDCWPNCVTIFNNGNHEAVKSMFHTQMRIELRALG
KISEGEELTVSYIDFLNVSEERKRQLKKQYFFDCTCEHCQKKLKDDLFLGVKDNPKPSQE
VVKEMIQFSKDTLEKIDKARSEGLYHEVVKLCRECLEKQEPVFADTNIYMLRMLSIVSEV
LSYLQAFEEASFYARRMVDGYMKLYHPNNAQLGMAVMRAGLTNWHAGNIEVGHGMICKAY
AILLVTHGSPHPITKDLEAMRVQTEMLRMFRQNEFMYYKMREAALNNQPMQVMAEPSNE
PSPALFHKKQ

A search of sequence databases reveals that the NOV62a amino acid sequence has 458 of 485 amino acid residues (94%) identical to, and 478 of 485 amino acid residues (98%) similar to, the 485 amino acid residue ptnr:SPTREMBL-ACC:P97443 protein from Mus musculus (Mouse) (ZINC-FINGER PROTEIN BOP). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV62a is expressed in at least Whole Organism, Heart, Lung, Prostate, Skeletal Muscle. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

NOV62b

A disclosed NOV62b nucleic acid of 1555 nucleotides (also referred to as CG57473-01) encoding a zinc finger BOP-like protein is shown in Table 62C. The start and stop codons are in bold letters.

Table 62C. NOV62b nucleotide sequence (SEQ ID NO:147).

AAGGTGCTGAAATAGCAATGACAAGAGACTTAGCTCAGTGTTAAATAACTGCCGCGCTGG
CCTGACAGTTTCTGAGATGACAATAGGGAGAATGGAGAACGTGGAGGTCTTCACCGCTGA
GGGCAAAGGAAGGGGTCTGAAGGCCACCAAGGAGTTCTGGGCTGCAGATATCATCTTTGC
TGATCGGGCTTATTCCGCAGTGTTTGTGACAGCCTTGTTAATTTTGTGTGCCACACCTG
CTTCAAGAGGCAGGAGAAGCTCCATCGCTGTGGGCAGTGCAAGTTTGCCCATTA
CCGACCTGCCAGAAGGATGCTTGGCTGAACCACAAGAATGAATGTTTCGGCCATCAAGAG
ATATGGGAAGGTGCCCAATGAGAACATCAGGCTGGCGGCGCGCATCATGTGGAGGGTGGA
GAGAGAAGGCACCGGGCTCACGGAGGGCTGCCTGGTGTCCGTGGACGACTTGCAGAACCA
CGTGGAGCACTTTGGGGAGGAGGAGCAGAAGGACCTGCGGGTGGACGTGGACACATTCTT
GCAGTACTGGCCGCCGAGAGCCAGCAGTTCAGCATGCAGTACATCTCGCACATCTTCGG
AGTGATTAAGTGAACGGTTTACTCTCAGTGATCAGAGAGGCCTGCAGGCCGTGGGCGT
AGGCATCTTCCCCAACCTGGGCCTGGTGAACCATGACTGTTGGCCCACTGTACTGT
ATTTAACAATGGCAATCATGAGGCAGTGAATCCATGTTTCATACCCAGATGAGAATTGA
GCTCCGGGCCCTAGGCAAGATCTCAGAAGGAGAGGAGCTGACTGTGTCTATATTGACTT
CCTCAACGTTAGTGAAGAACGCAAGAGGCAGCTGAAGAAGCAGTACTACTTTGACTGCAC
ATGTGAACACTGCCAGAAAAAACTGAAGGATGACCTCTTCTGGGGGTGAAAGACAACCC
CAAGCCCTCTCAGGAAGTGGTGAAGGAGATGATACAATTCTCCAAGGATACATTGGA
GATAGACAAGGCTCGTTCCGAGGGTTTGTATCATGAGGTTGTGAAATTATGCCGGGAGTG
CCTGGAGAAGCAGGAGCCAGTGTGTGCTGACACCAACATCTACATGCTGCGGATGCTGAG
CATTGTTTTCGAGGTCTTTTCTACCTCCAGGCCTTTGAGGAGGCCTCGTTCTATGCCAG
GAGGATGGTGGACGGCTATATGAAGCTCTACCACCCCAACAATGCCCAACTGGGCATGGT
CGTGATGCGGGCAGGGCTGACCAACTGGCATGCTGGTAACATTGAGGTGGGGCACGGGAT
GATCTGCAAAGCCTATGCCATTCTCCTGGTGACACACGGACCCCTCCACCCCATCACTAA
GGACTTAGAGGCCATGCGGGTGCAGACGGAGATGGAGCTACGCATGTTCCGCCAGAACGA

ATTCATGTACTACAAGATGCGCGAGGCTGCCCTGAACAACCAGCCCATGCAGGTCATGGC
CGAGCCCAGCAATGAGCCATCCCCAGCTCTGTTCCACAAGAAGCAATGAGGACTG

In a search of public sequence databases, the NOV62b nucleic acid sequence, located on chromosome 2 has has 1356 of 1525 bases (88%) identical to a gb:GENBANK-ID:MMU76373|acc:U76373.2 mRNA from Mus musculus (Mus musculus skm-BOP1 (Bop) mRNA, complete cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV62b polypeptide (SEQ ID NO:148) encoded by SEQ ID NO:147 has 490 amino acid residues and is presented in Table 62D using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV62b has no signal peptide and is likely to be localized in the cytoplasm with a certainty of 0.6500.

Table 62D. Encoded NOV62b protein sequence (SEQ ID NO:148).

MTIGRMENVEVFTEAGKGRGLKATKEFWAADIIFADRAYSAVVFDSLNVFVCHTCFKRQE
KLHRCGQCKFAHYCDRTCQKDAWLNHKNECSAISKRYGKVPNENIRLAARIMWRVEREGTG
LTEGCLVSVDDLQNHVEHFGEEEQKDLRVDDVDTFLQYWPPQSQQFSMQYISHIFGVINCN
GFTLSDQRGLQAVGVGIFPNLGLVNHDCWPNCVTIFNNGNHEAVKSMFHTQMRIELRALG
KISEGEELTVSYIDFLNVSEERKRQLKKQYYFDCTCEHCQKKLDDLFLGVKDNPKPSQE
VVKEMIQFSKDTLEKIDKARSEGLYHEVVKLCRECLEKQEPVFADTNIYMLRMLSIVSEV
LSYLQAFEEASFYARRMVDGYMKLYHPNNAQLGMVVMRAGLTNWHAGNIEVGHGMICKAY
AILLVTHGPSHPITKDLEAMRVQTEMELRMFRQNEFMYKMR EAALNNQPMQVMAEPSNE
PSPALFHKKQ

A search of sequence databases reveals that the NOV62b amino acid sequence has 457 of 485 amino acid residues (94%) identical to, and 478 of 485 amino acid residues (98%) similar to, the 485 amino acid residue ptrn:SPTREMBL-ACC:P97443 protein from Mus musculus (Mouse) (ZINC-FINGER PROTEIN BOP). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV62b is expressed in at least Whole Organism, Heart, Lung, Prostate, Skeletal and Muscle. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV62b polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 62E .

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 5870832 gb AAC53021.2 (U76373)	skm-BOP1 [Mus musculus]	485	458/485 (94%)	478/485 (98%)	0.0
gi 10257425 ref NP033892.1 (NM_009762)	CD8beta opposite strand [Mus musculus]	472	444/485 (91%)	465/485 (95%)	0.0
gi 1809322 gb AAC53020.1 (U76371)	t-BOP [Mus musculus]	456	419/447 (93%)	437/447 (97%)	0.0
gi 16930387 gb AAL31880.1 AF410781.1 (AF410781)	cardiac and skeletal muscle-specific BOP1 [Gallus gallus]	486	397/486 (81%)	441/486 (90%)	0.0
gi 16930389 gb AAL31881.1 AF410782.1 (AF410782)	cardiac and skeletal muscle-specific BOP2 [Gallus gallus]	473	384/486 (79%)	428/486 (88%)	0.0

Table 62F-G lists the domain descriptions from DOMAIN analysis results against NOV62b. This indicates that the NOV62b sequence has properties similar to those of other proteins known to contain this domain.

5

<p>Table 62F. Domain Analysis of NOV62b</p> <p>gnl Pfam pfam01753, zf-MYND, MYND finger.</p> <p>CD-Length = 38 residues, 100.0% aligned</p> <p>Score = 57.8 bits (138), Expect = 1e-09</p>

<p>Table 62G. Domain Analysis of NOV62b</p> <p>gnl Smart smart00317, SET, SET (Su(var)3-9, Enhancer-of-zeste, Trithorax) domain; Putative methyl transferase, based on outlier plant homologues</p> <p>CD-Length = 125 residues, 44.0% aligned</p> <p>Score = 53.9 bits (128), Expect = 2e-08</p>
--

Transcriptional regulatory proteins containing tandemly repeated zinc finger domains are thought to be involved in both normal and abnormal cellular proliferation and differentiation. One abundant class of such transcriptional regulators resembles the Drosophila Kruppel segmentation gene product due to the presence of repeated Cys2-His2 (C2H2) zinc

10

finger domains that are connected by conserved sequences, called H/C links. See ZNF91 (603971) for general information on zinc finger proteins.

By screening a human insulinoma cDNA library with a degenerate oligonucleotide corresponding to the H/C linker sequence, Tommerup et al. (1993) isolated cDNAs potentially encoding zinc finger proteins. Tommerup and Vissing (1995) performed sequence analysis on a number of these cDNAs and identified several novel zinc finger protein genes, including ZNF36, which they called ZNF139. The ZNF139 cDNA predicts a protein belonging to the Kruppel family of zinc finger proteins.

By isotopic in situ hybridization, Rousseau-Merck et al. (1995) mapped the ZNF36 gene, which they called KOX18, to 7q21-q22. From pulsed field gel electrophoresis studies, they showed that KOX18 is within less than 250 kb of KOX25 (ZNF38; 601261). Rousseau-Merck et al. (1995) tabulated 18 different KOX genes that had been located in pairs within 9 DNA fragments of 200 to 580 kb on 7 different chromosomes. By FISH, Tommerup and Vissing (1995) mapped the ZNF36 gene to 7q21.3-q22.1.

The disclosed NOV62 nucleic acid of the invention encoding a zinc finger BOP-like protein includes the nucleic acid whose sequence is provided in Table 62A or 62C or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 62A or 62C while still encoding a protein that maintains its zinc finger BOP-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 12 percent of the bases may be so changed.

The disclosed NOV62 protein of the invention includes the zinc finger BOP-like protein whose sequence is provided in Table 62B or 62D. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 62B or 62D while still encoding a protein that maintains its zinc finger

BOP-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 6 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

5 The above defined information for this invention suggests that this zinc finger BOP-like protein (NOV62) may function as a member of a “zinc finger BOP family”. Therefore, the NOV62 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to:

10 protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

15 The NOV62 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the zinc finger BOP-like protein (NOV62) may be useful in gene therapy, and the zinc finger BOP-like protein (NOV62) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for

20 treatment of patients suffering from Cardiomyopathy, Atherosclerosis, Hypertension, Congenital heart defects, Aortic stenosis , Atrial septal defect (ASD), Atrioventricular (A-V) canal defect, Ductus arteriosus , Pulmonary stenosis , Subaortic stenosis, Ventricular septal defect (VSD), valve diseases, Tuberous sclerosis, Scleroderma, Obesity, Transplantation, Aneurysm, Fibromuscular dysplasia, Stroke, Anemia , Bleeding disorders,

25 Adrenoleukodystrophy , Congenital Adrenal Hyperplasia, Diabetes, Von Hippel-Lindau (VHL) syndrome , Pancreatitis, Hyperparathyroidism, Hypoparathyroidism, SIDS, Endometriosis, Fertility, Xerostomia, Hypercalcaemia, Ulcers, Cirrhosis, Inflammatory bowel disease, Diverticular disease, Hirschsprung's disease , Crohn's Disease, Appendicitis, Hemophilia, hypercoagulation, Idiopathic thrombocytopenic purpura, autoimmune disease,

30 allergies, immunodeficiencies, Graft versus host, Ataxia-telangiectasia, Hemophilia, Lymphedema, Tonsillitis, Osteoporosis, Arthritis, Ankylosing spondylitis, Scoliosis, Tendinitis, Muscular dystrophy, Lesch-Nyhan syndrome, Myasthenia gravis, Dental disease and infection, Alzheimer's disease, Tuberous sclerosis, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-

TTCCTAACTCATTTTATGAGGCAAGCATCATGCTGATGCCAAATCTGGCAGAGACACAACAAAAAAGA
AAATTTTCAGGCCTATATCCCTGATGAACATCGATGTGAAAATCCTCAATAAAATACTGGCAAACCAATC
TTGCAGCACATCAAAAAGCTTATCCACGATGATCAAGTTGGCTTCATCCCTGGGATGCAAGGCTGGTTCA
ACATATGCAATCAATCAACATAATCCATCACATAAATAGCACCAATGACAAAAACCACATGATTATCTC
AATAGATGCGAAAAAGGCTTTGGTAAATTCACACCCCTTCATGCTAAAAACTTAAATAAGCTAGGT
ATTGATGGAACTGATCTCAAAATAAAGAGCTGTTTATGACAAACCCACAGCCAATATCATACTGACTG
GGCAAAGCTGGAAGCATTCCCTTTGAAAACCAGCACAAAGACAAGTATGCCCTCTCTACCACTCCTATT
CAACATGGTATTGGAAGTTCTGGCTAGGGCAATCAGGCAAGAGAAAGAAATAAAGCATATCCAAATAGGA
AGAGAGGAAGTCAAATGTCCCTGTTTGCAGATGACATGATTGTATATTTAGAAAACCCCATCGTCTCAG
CCAAAATCTCCTTAAGCTGATAAGAACTTCAGCAAAGTCTCGGGATACAAAATCAATGTGCAAAAATC
ACAAGCATTCCTATACATCAATAATAGACAAACAGAGAGCCAAATCGTGAGTGAACCTCCATTCACAATT
GTTACAAAGAGAATACAATACCTAGGAATACAACCTTACAAGGATGTGAAGGACCTCTTCAAGGAGAACT
ACAAACCACTGCTCAAGGAAATAAGAGAGGACACAAACAAATGGAAAAACATTCTATGCTCATGGATAGG
AAGAATCAATATCGTGAAAATGACCATGCTGCCCAAAGTAATTTATAGATTCAACACTATGCCCATCAAG
CTACCATTTGACTTTCTTACGGAATCAGACAAAACCTACTTTAAATTTTCATATGGAACCAAAAAAGAGCCT
GCACAGCCAAGACAATCCTAAGCAAAAAGAACAAAGCTGGAGGCATCACACTACCTAACTTCAAACTATA
CTACAAGGCTACAGTGACCAAAACAGCATGGTACTGGTACCAAAACAGATATACAGACCAATGGAACAGA
ATAGAGGCCCTCAGAAATAACACCACACATCTACAACCACCTGATCTTTGACAAACCTGACACAAACAAGC
AATGGGGAAAAGGATCTCTATTTAATAAATGGTGTGGGAAAACCTGGCTAGCCATATGCAGAAAACCTGAA
ACTGGACCCCTTCTTACACTTTTATACAAAATTAATTCAAGCTGGATTAAAGACTTAAATGTAAGACCT
AAAACAATAAAAAATCCTAGAAGAAAACCTGGGCAATACCATTAGGACATAGGCATGGGCAAAGACTTCG
TGACTGTAAACACCAAAAGCAATGGCAACAAAAGCCAAATTGACAAATGGGATCTAATTAACTAAAGAG
CTTCTGCACAGCAAAAGAACTGTCTCAGGGTGAACAGGCAACCTACAGAATGGGAAAAATTTTTTGCA
ATCTGTCCATCTGACAAAGGGCTAATATCCAGAATCTACAAGGAACCTTAAACAAATTTACAAGAAAAA
CAAAACACCTATCAAAAAGTGGGCAAGGCTATGAACAGACACTTCTCAAAAGAAGACATTTATGCAGC
CAAAAGACATATGAAAAATGGTCATCATCACTGGTCTTCAGGGAAATGCAAATCAAAACCACAATGAGA
TACCATCTCATGCCAGTTAGAATGGTGATCATTAGAAAGTCAGGAAACAACATGCATGCAATCAAAA
CCACAATGAGATACCATCTCATGCCAGTTAGAATGGTGATCATTAGAAAGTCAGGAAACAACATGCAG
AGGATGTGGAGAAATAGGAATGCTTTTACACTGTTGGTGGGAGTGAACTAGTTCAACCATTTGTTGAAG
ACAGTGTGGCGATTCTCAAGGATCTAGAACCAGAAATACCATTAGACCCAGCAATCCATTACTGGGTA
TATACCCAAATGATTATAAATCATGCTACTATAAAGACACATGCACACGTATGTTTATGCGGCACATT
CACAATAGCAAAGACTTGAACCAACCCAAATGCCCATCAGTGAGAGTCATAAAGAAAAATGTGGCACATA
TACATCATGGAATACTATGCAGCCATAAAAAAGGATGAGTTATGTCCTTTGCAGGGACATGGATGAATC
TGAAACCACCATTTCTCAGCAAACTAACACAGGAACAGAAAACCAATACCGCTTGTCTCACTCGTAAG
TTGGAGTTGAACAATGAGAACACATGGACACAGGGAGGGGAACAACACCGGGCCTGTAGGGGGTAGGG
GGGATAGGGGAGGGATAGCATTAAGAGAAATACCTAATGTAGATGACGGGTGATGGGTGCAGCAACCA
CCAAGGC

In a search of public sequence databases, the NOV63 nucleic acid sequence, located on
chromosome 13 has 3146 of 3647 bases (86%) identical to a gb:GENBANK-
ID:HSIL25FL|acc:X67285.1 mRNA from Homo sapiens (H.sapiens gene for interleukin-2 (5'
5 flanking region)). Public nucleotide databases include all GenBank databases and the
GeneSeq patent database.

The disclosed NOV63 polypeptide (SEQ ID NO:150) encoded by SEQ ID NO:149 has
1081 amino acid residues and is presented in Table 63B using the one-letter amino acid code.
Signal P, Psort and/or Hydropathy results predict that NOV63 has no signal peptide and is
10 likely to be localized in the cytoplasm with a certainty of 0.6000.

Table 63B. Encoded NOV63 protein sequence (SEQ ID NO:150).
MEIITNSLSDHSAIKLELRRIKKLTQNHTTTWKLKNLLNNYLVNNEIKAEINKFCETNEN KDTTYQNFWDTAKAVVRGKFIALNAHRRKQERCKINTLTSQLEKEKQEQTNSKANRRQE ITKIIAELKEIKTRKTHQKINESGSWFFEKINKIDRQLARLIKKRREKNQIDAINKDKGD ITADPTEIQTFTIREYYKHLIYANKLENLEEMDKFLATCTLPRLNQEELESNLRQITSSEIK AVINSLPTKQKPGPDGFTAIFYQRYKEELVPFLLKLFQTEKEGLLPNSFYEASIMLMPK SGRDTTKENFRPISLMNIDVKILNKILANQILQHIKKLIHDDQVGFIPGMQGWFNICKS INI IHHINSTNDKNHMIISIDA EKAFGKI QHPFMLKTLNKL GIDGTYLKI IRAVYDKPTA

NIILTGQKLEAFPLKTSTRQVCPLSPLLFMVLEVLARAIRQEKEIKHIQIGREEVKLSL
FADDMIVYLENPIVSAQNLLKLIRNFSKVSQGYKINVQKSQAFLYINNRRQTESQIVSELPF
TIVTKRIQYLGILQTRDVKDLFKENYKPLLKEIREDTNKWNILCSWIGRINIVKMTMLP
KVIYRFNTMPIKPLPLTFFTESDKTTLNFIWNQKRACTAKTILSKKNKAGGITLPNFKLYY
KATVTKTAWYQYQNRITDQWNRIEASEITPHIYNHLIFDKPDTNKQWGKGSFLNKCWEN
WLAICRKLKLDPLTLYTKINSSWIKDLNVRPKTIKILEENLGNTIQDIGMGKDFVTVTP
KAMATKAKIDKWDLIKLSFCTAKETVIRVNRQPTWEKFFAICPSDKGLISRIYKELKQ
IYKKKTNNPIKKWAKAMNRHFSKEDIYAAKRHMKKWSSSLVFREMQIKTTMRYHLMVPRM
VIIRKSGNNTCMQIKTTMRYHLMVPRMVIIRKSGNNTCRGCGEIGMLLHCWWECKLVQPL
WKTVWRFLKDLEPEIPLDPAIPLLLGIYPNDYKSCYYKDTCTRMFIAALFTIAKTWNQPKC
PSVRVIKKMWHIYIMEYYAAIKKDEFMSFAGTWMNLETTILSKLTQEQTQKRYRFLSVSW
S

A search of sequence databases reveals that the NOV63 amino acid sequence has 752 of 865 amino acid residues (86%) identical to, and 787 of 865 amino acid residues (90%) similar to, the 1010 amino acid residue patp:B38012 Human secreted protein encoded by gene 3 clone HNHCT15. Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

The disclosed NOV63 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 63C.

Table 63C. BLAST results for NOV63					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 106322 pir B34087	protein (L1H 3' region) - human	1280	919/1078 (85%)	966/1078 (89%)	0.0
gi 2072948 gb AAC51261.1 (U93563)	putative p150 [Homo sapiens]	1275	904/1078 (83%)	959/1078 (88%)	0.0
gi 339777 gb AAB59368.1 (M80344)	ORF2 contains a reverse transcriptase domain. [Homo sapiens]	1275	898/1078 (83%)	957/1078 (88%)	0.0
gi 5052951 gb AAD38785.1 AF149422.2 (AF149422)	unknown [Homo sapiens]	1275	900/1078 (83%)	957/1078 (88%)	0.0
gi 2136112 pir S65824	reverse transcriptase homolog - human transposon L1.1	1275	898/1078 (83%)	957/1078 (88%)	0.0

Table 63D lists the domain descriptions from DOMAIN analysis results against NOV63. This indicates that the NOV63 sequence has properties similar to those of other proteins known to contain this domain.

Table 63D. Domain Analysis of NOV63

gnl|Pfam|pfam00078, rvt, Reverse transcriptase (RNA-dependent DNA polymerase). A reverse transcriptase gene is usually indicative of a mobile element such as a retrotransposon or retrovirus. Reverse transcriptases occur in a variety of mobile elements, including retrotransposons, retroviruses, group II introns, bacterial msDNAs, hepadnaviruses, and caulimoviruses.

CD-Length = 208 residues, 97.6% aligned

Score = 99.0 bits (245), Expect = 1e-21

Secreted proteins can act as cytokines, growth factors, chemotactic factors, and ligands for cell surface receptors. Secreted functions play vital roles in the regulation of cell motility, proliferation, differentiation and apoptosis.

5

The disclosed NOV63 nucleic acid of the invention encoding a secreted protein-like protein includes the nucleic acid whose sequence is provided in Table 63A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 63A while still encoding a protein that maintains its secreted protein-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 14 percent of the bases may be so changed.

10

15

20

The disclosed NOV63 protein of the invention includes the secreted protein-like protein whose sequence is provided in Table 63B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 63B while still encoding a protein that maintains its secreted protein-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 14 percent of the residues may be so changed.

25

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this secreted protein-like protein (NOV63) may function as a member of a “secreted protein family”. Therefore, the NOV63 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV63 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the secreted protein-like protein (NOV63) may be useful in gene therapy, and the secreted protein-like protein (NOV63) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from CNS disorders, brain disorders including epilepsy, eating disorders, schizophrenia, ADD; cancer; heart disease; inflammation and autoimmune disorders including Crohn’s disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, blood disorders; psoriasis colon cancer, leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, pancreatic disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer, or other pathologies or conditions. The NOV63 nucleic acid encoding the secreted protein-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV63 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV63 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV63 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV64

A disclosed NOV64 nucleic acid of 3081 nucleotides (also referred to as CG57779-01) encoding a secreted protein-like protein is shown in Table 64A. The start and stop codons are in bold letters.

5

Table 64A. NOV64 nucleotide sequence (SEQ ID NO:151).

AAATGTAAAGAAGTGTCTCTCAGACCATGGTGCAATCAAACTAGAACTCAGCATTAAAGAACTCACTC
 AAAACTGCCCACTACATGGAACTGAACAACCTGCTCCTGAATGACTACTGGGTACATAACAAATCAA
 GGCAGAAATAAAGATGTTCTTTGAAACCAACAAGAACAAAGACACAACATACCAGAATCTCTGGGACACA
 TTCAAAGCAGTGTGTAGAGGGAAATTTATAGCACTAAATACCCATAAGAGAAAGCAGGAAAGATCTAAAA
 TTGACACCCTAACATCACAATTAATAACAACCTACAGAAGCAAGAGCAAAACACATTCAAAAGCTAGCAGAAG
 GCAAGAAATAACTAAGATCAGAGCAGAAGTGAAGGAGATAGAGACACAAAAAATGCTTCAAAAAAAT
 GAACCTTAAAAAGATAAAGGGGTTTGGTCCACCGATCCAGAGAAAAACACACTACCATCAGAGAATACT
 ATAAACACCTGTATGCAATAAACTAGAAAATCTGGAAGAAATGGATAAATTTCTGGACAAATACACCTT
 CCCAAGACTAAACCAGGAAGAAGTTGAATCCCTGAATAGACCAATAACAGGCTCGGAAATTGAGGCAATA
 ATTAATAGCTTACCAACCAAAAAAGTCCAGGGTCAGATGGATTACAGCCGAATTTCTACCAGAGGTACA
 AGGAGGAGCTGGTACCATTCTTCTGAACTATTCCAATCAATAGAAAAAGAGGGAATCCTCCCTAATCT
 ATTTGATGAGGCCAGCATCATCCTGTATACCAAAGCCTAGCAGAGACACAACAAAAAGAGAATTTTAGA
 CCAATATCCCTGATGAACATCGATGCAAAAATCCTCAATAAAATACTGGCAAAACGAATCCAGCAGCACA
 TCAAAAAGTTTATCCACCACGATCAAGTGGGCTTATCCCTAGGATGCAAGGCTGGTTTACATATGCAA
 ATCAATAAACGTAATCCAGCATATAAATAGAACCAGCAAAAGACAAAAACCATGATTATCTCAATAGATGCA
 GAAAAGGCCTTTGACAAAATTCACAGCCCTTCATGCTAAAACTCTCAGTAAATTAGGTATTGATATGA
 CATATCTCAAAATAATAAGAGCTATCTATGACAAACCCACAGCCAATATCATACTGAATGGGCAAAACT
 GGAAGCATTCCCTTTGAAAACCTGGCACAAGACATGGGTGCCCTCTCTCACCACCTCTATTCAACATAGTG
 TTGGAAGTCTGGCCAGGGCAATCAGGCAGGAGAGGAAATAAAGGGTATCAATTAGGAAAAGAGGAAG
 TCAAAATGTCCCTGTTTGCAGATGACATGATTTTATATCTAGAAAACCCCATCGTCTCAGCCCCAAATCT
 CCTTAAGCTGATAAGCAACTTCAGCAAAAGTCCCAGGATACAAAATCAATGTGCAAAAAACACAAGCATTC
 TTATACACCAATAACAGACAGACAGAGAGCCAATCATGAGTGAAGTCCCATTTACAATTGCTTCAAAGA
 GAATAAAATACCTAGGAATCCAACCTTACAAGGATGTGAAGGACTCTTCAAGGAGAACTACAAACCCATG
 CTCAATTGAAATAAAGAGGATACAAACAAATGGAAGAACATTCATGCTCATGGGTAGAAGAATCAAT
 ATTTGTGAAAATGGCCATTCTGCCCAAGGTAATTTATAGGTTCAATGCCATCCCCATCAAGCTACCAATGG
 CTTTCTTACAGAATTGGAAAAAATCTTTTAAAGTTTATATGGAACCAAAAAAGAGCCTGCATTGTCTAA
 GCCTGCATTGCTAAGCCAAAAGAACAAAGCTGGAGGCATCATGCTACCTGACTTCAAACATACTACAAG
 GCCACAGTAACCAAAACAGCATGGTACTGGTACCAAAACAGATATATAGACCAATGGAACAAAGCAGAGC
 CCTCAGAAATAATGCCACACATCTATAACTATCTGATCTTTGACAAACCTGACAAAAACAAGAAATCGGG
 AAAGGATTCCGTATTTAATAAACGGTCTGGGAAAACCTGGCTAGCCATATGTAGAAAGCTGAAACTGGAC
 CCTTCTCTTACACCTCATACAAAATTAATTCAGATGGATTAAAGACTTAAATGTTAGACTTAAACCA
 TAAAAACCTTAGAAGAAAACCTAGGCAATACCATTCAAGACATAGGCATGGGCAAGGACTTCATGTCTAA
 AACACCAAAAGCAATGGCAACAAAAGACAAAATTGACAAATGGGATCTAATTAACTAAAGAGCTTCTGC
 ACAGCAATAGAAATACCATCAGAGTGAACAGGCAACCTACAGAATGGGAGAAAAATTTTGCACCTACT
 CATCTGACAAAGGGCTAATATCCAGAATCCACAATGAACCTCAACAAATTTACAAGAAAAAATCAACAA
 CCCCATCAAAAAGTGGGCAAGGATATGAACAGACACTTCTCAAAAGAAGACATTTATGCAGCCAAAAGA
 CACATGAAAAAATGCTCATCATCTGGCCATCAGAGAAATGCAATGAAACACCAATGAGATACCATC
 TCACACCAGTTAGAATGGCGATCATTAATAAGTCAAGAAACACAGGTGCTGGAGAGGATGTGGAGAAAT
 AGGAACACTTTTACGCTGTTGGTGGGACTGTAACTAGTTCAACCATTTGTGGAAGTCAGTGTGGCGATT
 CTCAGGGATCTAGAACTAGAAATACCATTTGACCCAGCCATCCCATTACTGGGTATATACCCAAAGGACT
 ATAAATCATGCTGCTATAAAGACACATGCAGCCGTATGTTTGTGTCAGCACTATTCAACACAGCAAGAC
 TTGGAACCAACCAATGTCCAACATGATAGACTGGATTAAAGAAATGTGGCACATATACCCATGGAA
 TACTATGCAGCCACAAAAAAGGATGAGTTTCATGTCCTTTGCAGGGACATGGATGAAGCTGGAAACCA
 TCATTCTCAGCAAACTATCACAAGGACAGAAAACCAACACTGCATGTTCTCACTCATAGGTGGGAATTA
 G

In a search of public sequence databases, the NOV64 nucleic acid sequence, located on chromosome 13 has 2741 of 3050 bases (89%) identical to a gb:GENBANK-
 ID:HSU157D4|acc:Z68871.1 mRNA from Homo sapiens (Human DNA sequence from

cosmid U157D4, between markers DXS366 and DXS87 on chromosome X). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV64 polypeptide (SEQ ID NO:152) encoded by SEQ ID NO:151 has 1017 amino acid residues and is presented in Table 64B using the one-letter amino acid code.

- 5 Signal P, Psort and/or Hydropathy results predict that NOV64 has no signal peptide and is likely to be localized in the cytoplasm with a certainty of 0.3906.

Table 64B. Encoded NOV64 protein sequence (SEQ ID NO:152).	
MVQSKLELSIKKLTQNCPTTWKLNLLNDYWVHNKIKAEIKMFFETNKNKDDTTYQNLWD TFKAVCRGKFIALNTHKRKQERSKIDTLTSQKQLQKQEQTHSKASRRQBITKIRAEKE IETQKNASKKNELKKDKGVWSTDPREKHTTIREYYKHLKYANKLENLEEMDKFLDKYTFPR LNQEVEESLNRPIITGSEIEAIINSLPTKKSPGSDGFTAEFYQRYKEELVPFLLKLFQSIE KEGILPNSFDEASIIILIPKPSRDTTKKENFRPISLMNIDAKILNKILAKRIQQHIKKFIH HDQVGFIPRMQGWFNICKSINVIQHINRTKDKNHMISIDAFAFDKIQQPFMLKTLTKL GIDMTYLIKIRAIYDKPTANIILNGQKLEAFPLKTGTRHGCPLSPLLFNIVLEVLARAIR QEKEIKGIQLGKEEVKLSLFADDMILYLENPIVSAQNLLKLISNFSKVPGYKINVQKSQA FLYTNNRQTESQIMSELPFTIASKRIKYLGIQLTRDVKDSSRRTNPNCSIEIKEDTNKWK NIPCSWVRRINIVKMAILPKVIYRFNAIPIKLPMAFFTELEKTTLKFIWNQKRACIAKPA LLSQKNKAGGIMLPDFKLYYKATVTKTAWYWYQNRIDQWNKAEPSEIMPHIYNYLIFDK PDKNKKSGKDSVFNKRSWENWLAICRKLKLDPLTPHTKINSRWIKDLNVRPKTIKTLKEE NLGNTIQDIGMGKDFMSKTPKAMATKDKIDKWDLIKLSFCTAIEETTIRVNRQPTWEKI FATYSSDKGLISRIHNELKQIYKKKSNNPIKKWAKDMNRHFSKEDIYAAKRHMKKCSSSL AIREMQMKTMTMYHLTPVRMAIIKKSGNNRCWRGCGEIGTLLRCWWDCKLVQPLWKSVMR FLRDLELEIPFDPAIPLGLIYPKDYKSCCYKDTCSRMFVAALFTTAKTWNQPKCPTMIDW IKKMWHIYTMEYYAATKKKDEFMSFAGTWMKLETIILSKLSQGQKTKHCMFSLIGGN	

- 10 A search of sequence databases reveals that the NOV64 amino acid sequence has 829 of 977 amino acid residues (84%) identical to, and 864 of 977 amino acid residues (88%) similar to, the 1010 amino acid residue patp:B38012 Human secreted protein encoded by gene 3 clone HNHCT15. Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

The disclosed NOV64 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 64C.

15

Table 64C. BLAST results for NOV64					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 106322 pir B34087	protein (L1H 3' region) - human	1280	929/1044 (88%)	956/1044 (90%)	0.0
gi 5052951 gb AAD38785.1 AF149422.2 (AF149422)	unknown [Homo sapiens]	1275	920/1044 (88%)	949/1044 (90%)	0.0
gi 2072948 gb AAC51261.1 (U93563)	putative p150 [Homo sapiens]	1275	919/1044 (88%)	946/1044 (90%)	0.0
gi 2072958 gb AAC51267.1 (U93567)	p150 [Homo sapiens]	1275	918/1044 (87%)	947/1044 (89%)	0.0

gi 5070622 gb AAD39 215.1 AF148856.2 (AF148856)	unknown [Homo sapiens]	1275	918/1044 (87%)	949/1044 (89%)	0.0
---	---------------------------	------	-------------------	-------------------	-----

Table 64D lists the domain descriptions from DOMAIN analysis results against NOV64. This indicates that the NOV64 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 64D. Domain Analysis of NOV64
gnl Pfam pfam00078, rvt, Reverse transcriptase (RNA-dependent DNA polymerase). A reverse transcriptase gene is usually indicative of a mobile element such as a retrotransposon or retrovirus. Reverse transcriptases occur in a variety of mobile elements, including retrotransposons, retroviruses, group II introns, bacterial msDNAs, hepadnaviruses, and caulimoviruses.
CD-Length = 208 residues, 97.6% aligned
Score = 109 bits (272), Expect = 9e-25

The disclosed NOV64 nucleic acid of the invention encoding a secreted protein-like protein includes the nucleic acid whose sequence is provided in Table 64A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 64A while still encoding a protein that maintains its secreted protein-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 11 percent of the bases may be so changed.

The disclosed NOV64 protein of the invention includes the secreted protein-like protein whose sequence is provided in Table 64B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 64B while still encoding a protein that maintains its secreted protein-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 16 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this secreted protein-like protein (NOV64) may function as a member of a "secreted protein family". Therefore, the NOV64 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV64 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the secreted protein-like protein (NOV64) may be useful in gene therapy, and the secreted protein-like protein (NOV64) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from CNS disorders, brain disorders including epilepsy, eating disorders, schizophrenia, ADD; cancer; heart disease; inflammation and autoimmune disorders including Crohn's disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, blood disorders; psoriasis colon cancer, leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, pancreatic disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer, or other pathologies or conditions. The NOV64 nucleic acid encoding the secreted protein-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV64 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV64 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV64 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in

understanding of pathology of the disease and development of new drug targets for various disorders.

NOV65

- 5 A disclosed NOV65 nucleic acid of 3021 nucleotides (also referred to as CG57781-01) encoding a secreted protein-like protein is shown in Table 65A. The start and stop codons are in bold letters.

Table 65A. NOV65 nucleotide sequence (SEQ ID NO:153).
AATGACTACTGAGTAAATAATGAAATGAAGGCAGAAATAAAGATGTTCTTTGAAACCAATGAGAACAAAG ACACAATGTACCAGAATCTCTGGGACACATTTAAAGCAGTGTGTAGAGGGAAATTTATAGCACTAAATGC CCACAAGAGAAAGCAGGAAAGATCTAAATCAACATCCTAACATCAGAGTTAAAAGAACTAGGGAAGCAA GAACAAACAAATTCAAAAGCTAGCAGAAGGCAGAAATAAATAAGATCAGAGCAGAACTGAAGGAGATAG AGACACAAAAAACCTTCAAAAAATCAATGAATCCAGGAGCTGGTTTTTTGAAAAGATCAACAAAAATTGA TAGACAACCTAGCAAGACCAATAAAGAAGAAAAGAGAGAAGAATCAAATAGATGCAACAAAAAATGATAAA GGGGATATCACCCTGATCCCACAGAAATACAACTACCATCAGAGAATACTATCAACACTTCTATGCAA ATATACTAGAAAAATCTAGAAGAAATGGATAAAATTCCTGGACACATACACTCTCCCAAGACTAAACCAGGA AGAAGTTGAATCTCTGTATAGACCAATAACAGGTTCTGAAATTTGAGGCAATAATTAATAGGCTACCAACC AAAAAAGTCCAGGACCAGATGGATTACAGCTGAATTTCTACCAGAGGTACAAAGAGGAGCTGGTACCAT TCCTTCTGAACTATTTTCAAGACAACAGAAAAAGAGGGACTCCTCCCTAATCTATTTATGAGGCCAGCAT CATCCTGACACCAAAACCTGGTAGAGACACAACAAAAAAGAGAATTTTATGCCAATATCCCTGATGAAC ATTGATGCCGAAAGTCTCAATAAAATACTGGCAAAAGCTTATCCACCACATCAAAAGCTTATCCACCACG GTCAACTTGGCTTCATCCCTGGGATGCAAGGCTGGTTCAACATATGCAAATCAATAAATGTAGTTTCATCA CATAAACAGAACCAATGACAAAAACCATGATTTATCTCAATAGATGCAGAAAAGGCCTTCGACAAATATT CAACACCACTTCATGCTAAAACTCTGAGTAACTAGGTATCGATGGAACATATCTGAAAAATAAAGAG CTATTTTATGACAAACCCACAGCCCAATATCATAGTGAATGGGCAAAACTGGAAGCATTCCTTTTGAAAAC TGGCACAAGACAAGGATGCCCTCTCTCACCCTCCTATTCAACATAGTGTGGAGTTCTGGCTAGGGCA ATCAGGCAAGAGAAAGAAATAAACGGTATTCAATTAGGAAAAGAGGAAGTCAAATGTCTCTGTGTGCAG ATGACATGATTGTATATTTAGAAAACCCCATCGTCTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTT CAGCAAAGTCTCAGGATACAAAATCAATGTGCAAAAATCACAAGCATTCCTATACATCAATAATAGACAA ACAGAGAGCCAAATCATGAGTGAATCCCATTTCCCAATTACCAAGAGAAATTAATACCTAGGAATCC AACTTACAAGGGATGTGAAAGACCTCTTCAAGGAGAACTACAAACCACTGCTCGAAATAAAAGAGGACAC AAACAAATGGAAAAACATTCCATGCTCATGGATAGGAAGAATCAATATTGTGAAAATGGTCATCTGCCC AAAGTAATTTATAGATTCAATGCTATCCCCATCAAGCTACCACTGACTTTCTTCACAGAATTGGAAAAAA CTATTTTAAAGTTTCATATGGAACCAAAAAAGAACCCAGATTGCCAAGACAATCCTAAGCAAAAAGAACAA AGCTGGAGGCATCACACTACCTGACTTCAAACTATACACAGGCTACAGTAAACAAAACAGCATGGGTAC TGGTACCAAAAACAGATATATAGACCAATGGAACAGAATGGAGGCCTCAGAAATAACACCACACATCTACA ACCATCTGATCTTTGACAAACCTGACAAAAACAGGCAATGGGGAAAGGATTCTCTATTTAATAAATGGTG CTGGGAAAACCTGGCTAGCCATATGTAGAAAGCTGAACTGGACCCCTTCTTACACCTTATACAAAAATT AACACAAGATGGATTAAAGACTTAAACGTCAGACCTAATACCATAAAAACCTAGAAAGAAAACCTAGGCA ATACCATTACAGGACATAGGCATGGGCAAGTCTTCATGACTAAAACACCAAAAGCAATGGCAACAAAAGT CAAAATTGACAAATGGGATCTAATTAACTAAAGAGCTTCTGCACAGCAAAAGAACTATCATCAGAGTG AACAGGCAACCTACAGAATGGGAGAAAATCTTTGCAACCTACCCATCTGACAAAGGCTAATATCCAGAA TCTACAAAGAACTCAACAAATTTACAAGAAAAAACAACCCCATCAAAAAGTGGGCAAAATACAAGA AAAAAAAACAACCCCATCAAAAAGTGGGCAAGGATATGAGCAGACACTTCTCAAAAGAAGACATTTAT GCAGCCAACAGAATGAAAAAGTGGTTCATCATCACTGGTCTTCAGGGAAATGCAAATCAAAACCACAATGA GATACCATCTCATGCCAGTTAGAAATGGTGATCATTAAAAAGTCAGGAAACAACACATGCCTGAGAGGATG TGGAGAAATAGGAATGCTTTTACACTGTTGGTGGGAGTGTAACCTAGTTCAACCATTTGTGGAAGACAGTG TGGCGATTCTCAAGGATCTAGAACCAAGAAATACCATTAGACCCAGCAATCCATTACTGGGTATATACC CAACCAATTATAATCATGCTACTATAAAGACACACGCACAGTATGTTTATTGTGGCAGTATTACGAT AGCAAAGAAGGATGAGTTTCATGCTCTTTGCAAGGACATGGATGAAGCTGGAACCATCATTTCTAAGCAA CTATCACAAGGACAGAAAACCAACACCACATGTTCTCACTCATAGGTGGGAGTTGAACAACGAGAACAC ATGGACACAGG

- 10 In a search of public sequence databases, the NOV65 nucleic acid sequence, located on chromosome 13 has 2524 of 2878 bases (87%) identical to a gb:GENBANK-

ID:F229117S02|acc:AF229118.1 mRNA from Homo sapiens (Homo sapiens acetylcholinesterase collagen-like tail subunit (COLQ) gene, exons 1A, 2, 3, 4, and 5). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV65 polypeptide (SEQ ID NO:154) encoded by SEQ ID NO:153 has 990 amino acid residues and is presented in Table 65B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV65 has no signal peptide and is likely to be localized in the cytoplasm with a certainty of 0.7000.

Table 65B. Encoded NOV65 protein sequence (SEQ ID NO:154).

MKAEIKMFFETNENKDTMYQNLWDTFKAVCRGKFIALNAHKRKQERSKINILTSQKELG
KQEQTNSKASRRQKITKIRAELEKEIETQKTLOKINESRSWFFKINKIDRQLARPIKKKR
EKNQIDATKNDKGDITTDPTETIQTIREYYQHFYANILEEMDKFLDTYTLPRNLQEE
VESLYRPITGSEIEAIINRLPTKKSPGPDGFTAIFYQRYKEELVPFLKLFQTTEKEGLL
PNSFYEASIIILTPKPGRDTTKKENFMPISLMNIDAKVLNKLAKAYPPHQKLIHHGQLGF
IPGMQGWFNICKSINVVHHINRTNDKNHMIISIDA EKAFDNIQHHFMLKTL SKLGIDGT
LKIIRAIYDKPTANII VNGQKLEAFPLKTGT RQCPLSPLLFNIVLEVLARAI RQEKEIN
GIQLGKEEVKLSLCADDMIVYLENP IVSAQNLLKLISNFSKVS GYKINVQKSQAFLYINN
RQTESQIMSELPFPITTKRIKYLGIQLTRDVKDLFKENYKPLLEIKEDTNKWKNI PCSWI
GRINIVKMVILPKVIYRFNAIPIKLP LTFTELEKTILKFIWNQKRTQIAKTILSKKNKA
GGITLPDFKLYYKATVNKTAWYWYQNR YIDQWNRMEASEITPHIYNHLIFDKPKDN RQWG
KDSL FNKWCWENWLAICRKLKLD PFLTPYTKINTRWIKDLNVRPNTIKTLEENLGNTIQD
IGMGKVFMTKTPKAMATKVKIDKWDLIKLSFCTAKETIIRVNRQPTWEKIFATYPSDK
GLISRIYKELKQIYKKKKTTSPSKSGQIQEKKNNPIKKWAKDMSRHF SKEDIYAANRMKKW
SSSLVFREM QIKTTMRYHLM PVRMVI IKKSGNNTCLRGCGEIGMLLHCWWECKLVQPLWK
TVWRFLKDLEPEIPLDPAIPL LGIYPNNYKSCYYKDTTRTRMFIVALFTIAKKDEFMSFAG
TWMKLETIILSKLSQGQKTKHHMFSLIGGS

A search of sequence databases reveals that the NOV65 amino acid sequence has have 842 of 951 amino acid residues (88%) identical to, and 867 of 951 amino acid residues (91%) similar to, the 1010 amino acid residue patp:B38012 Human secreted protein encoded by gene 3 clone HNHCT15. Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

The disclosed NOV65 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 65C.

Table 65C. BLAST results for NOV65

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 106322 pir B340 87	protein (L1H 3' region) - human	1280	885/1023 (86%)	911/1023 (88%)	0.0

gi 17980447 gb AAL50637.1 (AF421375)	unknown [Homo sapiens]	1275	870/1023 (85%)	904/1023 (88%)	0.0
gi 2072948 gb AAC51261.1 (U93563)	p150 [Homo sapiens]	1275	870/1023 (85%)	903/1023 (88%)	0.0
gi 2136112 pir S65824	reverse transcriptase homolog - human transposon L1.1	1275	868/1023 (84%)	904/1023 (87%)	0.0
gi 5052951 gb AAD38785.1 AF149422.2 (AF149422)	unknown [Homo sapiens]	1275	868/1023 (84%)	904/1023 (87%)	0.0

Table 65D lists the domain descriptions from DOMAIN analysis results against NOV65. This indicates that the NOV65 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 65D. Domain Analysis of NOV65
gnl Pfam pfam00078, rvt, Reverse transcriptase (RNA-dependent DNA polymerase). A reverse transcriptase gene is usually indicative of a mobile element such as a retrotransposon or retrovirus. Reverse transcriptases occur in a variety of mobile elements, including retrotransposons, retroviruses, group II introns, bacterial msDNAs, hepadnaviruses, and caulimoviruses.
CD-Length = 208 residues, 97.6% aligned
Score = 108 bits (269), Expect = 2e-24

The disclosed NOV65 nucleic acid of the invention encoding a secreted protein-like protein includes the nucleic acid whose sequence is provided in Table 65A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 65A while still encoding a protein that maintains its secreted protein-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 13 percent of the bases may be so changed.

The disclosed NOV65 protein of the invention includes the secreted protein-like protein whose sequence is provided in Table 65B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 65B while still encoding a protein that maintains its secreted protein-like activities
5 and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 12 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this secreted protein-
10 like protein (NOV65) may function as a member of a “secreted protein family”. Therefore, the NOV65 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to:
15 protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV65 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the secreted protein-like
20 protein (NOV65) may be useful in gene therapy, and the secreted protein-like protein (NOV65) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from CNS disorders, brain disorders including epilepsy, eating
25 disorders, schizophrenia, ADD; cancer; heart disease; inflammation and autoimmune disorders including Crohn’s disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, blood disorders; psoriasis colon cancer, leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, pancreatic disorders including pancreatic insufficiency and cancer; and
30 prostate disorders including prostate cancer, or other pathologies or conditions. The NOV65 nucleic acid encoding the secreted protein-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

antibodies that bind immuno-specifically to the novel NOV65 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV65 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV66

A disclosed NOV66 nucleic acid of 3120 nucleotides (also referred to as CG57783-01) encoding a secreted protein-like protein is shown in Table 66A. The start and stop codons are in bold letters.

Table 66A. NOV66 nucleotide sequence (SEQ ID NO:155).

CTGAATGACTACTGGGTACATAACAAAATGAAGACAGAAATAAAGATGTTCTTTGAAACCAATGAGAACA
AAGACACAACATACCAGAATCTCTGGGACACATTCAAAGCAGTGTGTAGAGGGAAATTTACAGCACTAAA
TGCCCATAGAGAAAGCAGGAAAGATCCAAAATTGACACCCTAACATCACAATTTAAACAACCTACAGAAG
CAAGAGCAAACACTTTCAAAGCTAGCAGAAGGCAAGAAATAACTAAGATCAGAGCAGAACTGAAGGAGA
TAGAGACACAAAAACCTTCAAATAATCAATGAATCCAGGAGCTGGTTTTTTGAAAAGATCAACAAAAT
TGATACACTGCTAGCAAGACTAATAAAGAAGAAAAGAGAGAAGAATCAAATAGACGCAATAAAAAATGAT
AAAGCAGATATCACCCTGATCCACAGAAATACAACTACCATCAGAGAATACTATAAACACCTCTATG
CAAATAAATAGAAAACTAGAAAGAAATGGATAAATTCCTTGACACATACACCTCCCAAGAAATAACCA
GGAAGAAGTTGAATCTCTGAATAGACCAATAACAGGCTCTGAAATTGAGGCAATAATTAATAGCTTACCA
ACCAAAAAAGTCCAGGACAGACGATTACAGCCGAATTTACCAGAAGTACAAGGAGGAGCTGATAC
CATTCCTTCTGAACTATTCCAATCAATAGAAAAGAGGGAATCCTCCCTAACTCATTGTGATGAGGCCAG
CATCATCTGATACCAAAGCCTGGCAGAGACACAACAAAAAGAGAATTTAGACCAATATCTCTGATG
AACATTTGATGCAAAATCCTCAATAAAATACTGGCAACCGAATCAAGCAACACATCAAAAAGCTTATCC
ACCATGATCAAGTGGGCTTCATCCTTGGGATGCAAGGCTGGTTCAACATATGCAATCAATAAACGTAAT
CCAGCATATAACAGAACCAAGACAAAAACCATGATTATCTCAATAGATGCAGAAAAGGCTTTGAC
AAAATTCAACAGCACTTCATGCTAAAACTCTCAATAAATTAGGTATTGATGGGACGTATCTCAAAATAA
TAAGAGCTATCTGTGACAAACCCACTGCCAATATCATACTGAATGGGCAAAAAGTGAAGCGTTCCCTTT
GAAAACCTGGCACAAGACAAGGGTGCCCTCTCTCACCCTCTATTCAACATAGTGTGGAGTCCCTGGCC
AGGGCAATCAGGCAGGAGAAGGAAATAAAGGGTATTGAGTTAGGAAAAGGGAAGTCAAATTGTCTCTGT
TTGCAGATGACATGATGTATATCTAGAAAACCCCATCATCTCAGCCCAAAATCTCCTTAAGCTGATAAG
CAACTTCAGCAAAGTCTCAGGATACAAAATCGATGTGCAAAAATCACAAGCATTCTTATACACCAATACA
GACCAGACAGAGGCAATCATGAGTGACCTCCCATTCACAATTGCTTCAAAGAGAATAAAATACCTAG
GAATCCAATTACAAGGGATGTGAAGGACCTCTCAAGGAGAACTACAACCACTGCTCAATGAAATAAA
AAAGGATACAAACAAATGGAAGAACATTCCAGGCTCATGGATAGGAAGAATCAATATCGTGAATGGCC
ATAGAGCCCAAGGTAATTTATAGATTCAATGCCATCCCCATCAAGCTACCAATGACTTTCTTACAGAAC
TGGAGAAAACCTTTAAAGTTCATATGGAACCAAAAGAGAGCCACATTGCCAAGTCAATCCTAAACCA
AAAGAACAAAGCTGGAGGCATCACACCCTGACTTCAAATATACTACAAGGCTACAGTAAACAAAACA
GCATGTTACTGGTACCAAAACAGAGATATAGACCAGTGGAAACAGACAGATCCCTCAGAAATAATGCCAC
ACATCTCAACTATCTGATCTTTGACAAACCTGACAAAAAGCAATGGGGAAAGGATTCCCTATTATA
TAAATGGTGTGGGAAAACCTGGCTAGCCATAGGTAGAAAGCTGAAACTGGACCCCTTCTTACACCTTAT
ACAAAAATTAATTCAAGATGGATTAAAGACTTAAATGTTAGACCTAAAACCATAAAAACCTTAGAAGGAA
ACCTAGGTATTACCATTTAGGACACAGGCATGGGCAAGGACTTCATGTCTAAAACACCAAAAGCAATGGC
AACAAAAGACAAAATTGACAAATGGGATCTAATTAACTAAAGAGCTTCTGCACAGCAAAAGAACTACC
ATCAGAGTGAACAGGCAACCTACAAAATGGGAGAACTTTTTCACACCTATTCTATCTGACAAAAGGCTAA
TATCCAGATCTACAAAGAACTCAACAAAATTTCAAGAAAAGAAACAAACACCCCATCAAAAAAAAAC
TAACAACCCCATCAAAAAGCGGGCAAGGATATGAACAGACACTTCTCAAAAGAAGACATTTATGCAGCC
AAAAGACACATGAAAAATGCTCATCATCACTGGCCATCAGAGAAATGCAATGAAAACCACAATGAGAT

ACCATCTCACACCAGTTAGAATGGCGATCATTAAAAAGTCAGGAAACAACAGGTGCTGGAGAGGATGTGG
 AGAAATAGGAACACTTTTACGCTGTTGGTGGGACTGTAAACTAGTTCAACCATGTGGAAGACAGTGTGG
 CGATTCTCAGGGATCTAGAAGTAAATACCATTTGACCCAGCCATCCCATTACTGGGTATATACCCAA
 AGGATTATAAATCATGCTGCTATAAAGACACATGCAGACGTATGTTTATTGCGGCACATTTACAATAGC
 AAAGACTTGGGAACCAACCCAAATGTCCAACAATGATAGACTGGATTAAGAAAATGTGGCACATATACACC
 ATGAAATACTATGCAGCCATAAAAAATGATGAGTTCATGTCTTTGTAGGGACATGGATGAAGCTGGAAA
 CCATCATCTCAGCAACTATCACAGGACAGAAAACCAACACCATGTTCTCACTCATAGGTGGAAA
 TTGAACAATGAGAATACTTTGACACAGGAAGGGGAACATC

In a search of public sequence databases, the NOV66 nucleic acid sequence, located on chromosome 13 has 2399 of 2567 bases (93%) identical to a gb:GENBANK-

ID:HSNOD1G2|acc:AF149774.1 mRNA from Homo sapiens (Homo sapiens NOD1 protein (NOD1) gene, exons 4 through 14 and complete cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV66 polypeptide (SEQ ID NO:156) encoded by SEQ ID NO:155 has 1018 amino acid residues and is presented in Table 66B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV66 has no signal peptide and is likely to be localized in the cytoplasm with a certainty of 0.6000

Table 66B. Encoded NOV66 protein sequence (SEQ ID NO:156).

MKTEIKMFFETNENKDTTYQNLWDTFKAVCRGKFTALNAHKRKQERSKIDTLTSQLKQLQ
 KQEQTLSKASRRQEITKIRAELEKEIETQKTLQKINESRSWFFFEKINKIDTLLARLIKKKR
 EKNQIDAIKNDKADITTDPTIEIQTIREYYKHLIYANKLENLEMDKFLDITYTLPRINQEE
 VESLNRPIITGSEIEAIINSLPTKKSPGPDGFTAIFYQKYKEELIPFLKLQFSIEKEGIL
 PNFDEASIIILPKPRDRTTKKENFRPISLNMIDAKILNKILANRIKQHIKKLIHHDQVG
 FIPGMQGWFNICKSINVIQHINRTKDKNHMIIISIDAEKAFDKIQQHFMKLTLNKLGLDGT
 YLKIIRAICDKPTANIILNGQKLEAFPLKTGTROGCPLSPLLNFIVLEVLARAIRQEKEI
 KGIQLGKEEVKLSLFADDMIVYLENPIISAQNLLKLISNFSKVSQYKIDVQKSQAFLYTN
 TDQTESQIMSDLPFTIASKRIKYLGIQLTRDVKDLFKENYKPLLEIKKDTNKWKNIPGS
 WIGRINIVKMAIEPKVIYRFNAIPIKLPMTFFTELEKTTLKFIWNQKRAHIAKSILNQKN
 KAGGITPPDFKLYYKATVNKTAWYWYQNRDIDQWNRDTPSEIMPHIYNYLIFDKPDKKKQ
 WGKDSLFLNKWCWENWLAIGRKLKLDPLTPYTKINSRWIKDLNVRPKTIKTLEGNLGI
 EDTGMGKDFMSKTPKAMATKDKIDKWDLIKLSFCTAKETTIRVNRQPTKWEKLFATYSS
 DKGLISRIYKELKQIYKKRTNNPIKKKTNNPIKKRAKDMNRHFSKEDIYAAKRMHMKCSS
 SLAIREMQMKTMMRYHLTPVRMAIIKKSNNRCWRGCGEIGTLLRCWWDCKLVQPLWKTV
 WRFLRDLELEIPFDPAIPLLGIIYKDYKSCCYKDTCTRMFIAALFTIAKTWNQPKCPTMI
 DWIKMWHIYTMKYAAIKNDEFMSFVGTWMKLETIIILSKLSQGQKTKHHMFSLIGCN

A search of sequence databases reveals that the NOV66 amino acid sequence has have 947 of 1018 amino acid residues (93%) identical to, and 965 of 1018 amino acid residues (94%) similar to, the 1010 amino acid residue patp:B38012 protein from human (Human secreted protein (L1H 3' region)). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

The disclosed NOV66 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 66C.

Table 66C. BLAST results for NOV66					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
<u>gi 106322 pir B34087</u>	protein (L1H 3' region) - human	1280	955/1018 (93%)	979/1018 (95%)	0.0
<u>gi 2072958 gb AAC51267.1 </u> (U93567)	putative p150 [Homo sapiens]	1275	945/1018 (92%)	971/1018 (94%)	0.0
<u>gi 5052951 gb AAD38785.1 AF149422.2</u> (AF149422)	unknown [Homo sapiens]	1275	948/1018 (93%)	972/1018 (95%)	0.0
<u>gi 5070622 gb AAD39215.1 AF148856.2</u> (AF148856)	unknown [Homo sapiens]	1275	945/1018 (92%)	973/1018 (94%)	0.0
<u>gi 2072953 gb AAC51264.1 </u> (U93565)	p150 [Homo sapiens]	1275	943/1018 (92%)	971/1018 (94%)	0.0

Table 66D lists the domain descriptions from DOMAIN analysis results against NOV66. This indicates that the NOV66 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 66D. Domain Analysis of NOV66
<u>gnl Pfam pfam00078</u> , rvt, Reverse transcriptase (RNA-dependent DNA polymerase). A reverse transcriptase gene is usually indicative of a mobile element such as a retrotransposon or retrovirus. Reverse transcriptases occur in a variety of mobile elements, including retrotransposons, retroviruses, group II introns, bacterial msDNAs, hepadnaviruses, and caulimoviruses.
CD-Length = 208 residues, 93.8% aligned
Score = 114 bits (285), Expect = 3e-26

The disclosed NOV66 nucleic acid of the invention encoding a secreted protein-like protein includes the nucleic acid whose sequence is provided in Table 66A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 66A while still encoding a protein that maintains its secreted protein-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical

stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 7 percent of the bases may be so changed.

5 The disclosed NOV66 protein of the invention includes the secreted protein-like protein whose sequence is provided in Table 66B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 66B while still encoding a protein that maintains its secreted protein-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 7 percent of the residues may be so changed.

10 The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this secreted protein-like protein (NOV66) may function as a member of a "secreted protein family". Therefore, the NOV66 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

20 The NOV66 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the secreted protein-like protein (NOV66) may be useful in gene therapy, and the secreted protein-like protein (NOV66) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from CNS disorders, brain disorders including epilepsy, eating disorders, schizophrenia, ADD; cancer; heart disease; inflammation and autoimmune disorders including Crohn's disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, blood disorders; psoriasis colon cancer, leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, pancreatic disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer, or other pathologies or conditions. The NOV66 nucleic acid encoding the secreted protein-like protein of the invention, or fragments thereof,

may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV66 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV66 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV66 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV67

NOV67 includes two acyltransferase-like proteins disclosed below. The disclosed sequences have been named NOV67a and NOV67b.

NOV67a

A disclosed NOV67a nucleic acid of 1116 nucleotides (also referred to as CG57823-01) encoding a acyltransferase-like protein is shown in Table 67A. The start and stop codons are in bold letters.

Table 67A. NOV67a nucleotide sequence (SEQ ID NO:157).

CCGGCACC GGCGTCAAGGCCATGGCGCTGTGCCCGGGATGGGTACACACCGAATTCCACT
CACGCGCCAACGTACCGGCAACCATCTGCCGGACTTTTCTGGATCGACGCCGAAGTTC
TGGTACGCGAGGCTCTCAACGACCTTGACCATGACAAGGTAGTATCCATTCTTACCCCGC
TCTGGAAGTTCTTCATCGCAGTGGCCACACATACCCACGTTCCGCTATGAGATTCTGT
CACGAACCTCTGTCCTCGTCTCGAGACAAGGACGACCATCCTCGACACACTCCGGGAGGCG
AGGCCTGAGATGGCCAGCGTCAAACCCACTAAGGACCGGGCCGGTACACCAATGATCTG
TCCGCGCGACGCGGCAGGCAGCGAACATGCTTCTGCTGCGTCTTTGGTGTGAAAGTC
GTCAAAGTGAGCGTCCACGGAGCCGACAACCTCGACGGGCTCGACGGTGCTTACGTCGCC
GTCGCTAACCATTCTCCACCTCGACGCGCCGCTCGTTTTTGGGGCCCTTCCAAGCGG
CTGTCAAAGTACCTAGCTACCGGGCCGCTGCTGACTATTTCTTACCGCCTGGTGGAAG
GCCATCGCTCCGGTGCTCTTCTCAACGCGTTCCCGGTCGACCGAGGCAAAGGCAAAGT
AAGCAAGGTGCCCGTAGTCCCGGTTCCACCGCGGTATGGCTGGGTCACTGCTGACAGAT
GGCGTCCCCCTGCTGATCTTTCCGGAGGGCACCCGGTCTCGCACCGGCGCAATGGGCACC
TTCAAACCTGGGGTGCCGATTGGCTATTTACGTGGGGTTCCGGTTATCCGATTGCT
TTAGTAGGACATGGGCGGCTATGCCGTCCGAGCAAGCCAGGTACCAAAAGGACGTCCA
TTGGTCCACGTGGCTATTGGACACCCTATGGACCCTGTTCCCGGCGAGATCGCCACCAA
TTCTCCGAACGGATTGCTCGCCAGGTCAATTGAGTTGCACGACCAAACCGCCCGGCCTAC
GGCATGCCAACCCTTGACGAATACGGACGCCACCGCGCGCTAAGCCAGGCCTCCGAGAGC
GGCGACACCGCATCCACCAACCACTCGACGTGAC

In a search of public sequence databases, the NOV67a nucleic acid sequence has 323 of 558 bases (57%) identical to a gb:GENBANK-ID:AF263912|acc:AF263912.1 mRNA from *Streptomyces noursei* (*Streptomyces noursei* ATCC 11455 nystatin biosynthetic gene cluster, complete sequence). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV67a polypeptide (SEQ ID NO:158) encoded by SEQ ID NO:157 has 267 amino acid residues and is presented in Table 67B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV67a has a signal peptide and is likely to be localized at the ER plasma membrane with a certainty of 0.85000. The most likely cleavage site for a NOV67a peptide is between amino acids 44 and 45.

Table 67B. Encoded NOV67a protein sequence (SEQ ID NO:158).
MASVKPTKDRGRYTNDLSAATRQAANMLLLRPLVWKVVKVSVHGADNLDGLDGAYVAVAN HSSHLDAPLVFGALPKRLSKYLATGAAADYFFTAWWKAIAFVLFNAFPVDRGKSKQGG ARSPRSHRGMAGSLTLDGVPLLI FPEGTRSRGTAMGTFKPGAAALAI SRGVPVPIPIALVG AWAAMPSEQARLPKGRPLVHVAIGHPMDFVPGEIAHQFSEIRIRQVIELHDQTARAYGMP TLDEYGRHRALSQASESGDTASTNHST

A search of sequence databases reveals that the NOV67a amino acid sequence has 65 of 181 amino acid residues (35%) identical to, and 96 of 181 amino acid residues (53%) similar to, the 240 amino acid residue ptrn:TREMBLNEW-ACC:CAC01452 protein from *Streptomyces coelicolor* (PUTATIVE ACYLTRANSFERASE). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV67a is expressed in at least Bone, Bone Marrow, Brain, Liver, Lung, Lymph node, Placenta, Prostate, Thalamus, Thyroid and Uterus. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

NOV67b

A disclosed NOV67b nucleic acid of 906 nucleotides (also referred to as CG57823-02) encoding a acyltransferase-like protein is shown in Table 67C. The start and stop codons are in bold letters.

Table 67C. NOV67b nucleotide sequence (SEQ ID NO:159).
ATACCCACGTTCCGCTATGAGATTCTGTACGAACTCTGTCTCGTCTCGAGACAAGG ACGACCAACCCTCGACACACTCCGGGAGGCGAGGCCTGAGATGGCCAGCGTCAAACCCACT AAGGACCGGGGCCGGTACACCAATGATCTGTCCGCCGCGACGCGGCAGGCAGCGAACATG

CCTCTGCTGCGTCCTTTGGTGTGGAAAGTCGTCAAAGTGAGCGTCCACGGAGCCGACAAC
CTCGACGGGCTCGACGGTGCCTACGTGCGCGTCGCTAACCATTCCTCCCACCTCGACGCG
CCGCTCGTTTTTGGGGCCCTTCCCAAGCGGCTGTCAAAGTACCTAGCTACCGGGGCCGCT
GCTGACTATTTCTTCACCGCCTGGTGGAGGCCATCGCTCCGGTGCTCTTCTTCAACGCG
TTCCCGGTGACCGAGGCAAAGGCAAAGTAAGCAAGGTGCCCGTAGTCCCCGTTCCAC
CGCGGTATGGCTGGGTCACTGCTGACAGATGGCGTCCCCCTGCTGATCTTCCGGAGGGC
ACCCGGTCTCGCACCGGTGCAATGGGCACCTTCAAACCTGGGGCTGCCGATTGGCTATT
TCACGTGGGGTTCCGGTTATCCCGATTGCTTTAGTAGGAGCATGGGCGGCTATGCCGTCC
GAGCAAGCCGGGTTACCAAAGGACGCCCATCGTCCACGTGGCTATTGGACACCCATG
GACCCTGTTCCCGGCGAGATCGCCACCAATTCTCCGAACGGATTGTCGCCAGGTCATT
GAGTTGCACGACCAAACCGCCCGCGCTACGGCATGCCAACCTTGACGAATACGGACGC
CACCGCGCGCTAAGCCAGGCCTCCGAGAGCGGCGACACCGCATCCACCAACCACTCGACG
TGACAC

In a search of public sequence databases, the NOV nucleic acid sequence has 323 of 558 bases (57%) identical to a gb:GENBANK-ID:AF263912|acc:AF263912.1 mRNA from *Streptomyces noursei* (*Streptomyces noursei* ATCC 11455 nystatin biosynthetic gene cluster, complete sequence). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV67b polypeptide (SEQ ID NO:160) encoded by SEQ ID NO:159 has 267 amino acid residues and is presented in Table 67D using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV67b has a signal peptide and is likely to be localized at the ER plasma membrane with a certainty of 0.85000. The most likely cleavage site for a NOV67b peptide is between amino acids 44 and 45.

Table 67D. Encoded NOV67b protein sequence (SEQ ID NO:160).

MASVKPTKDRGRYTNDLSAATRQANMLLLRPLVWKVSVHAGDNLGDLGAYVAVAN
HSSHLDAPLVFGALPKRLSKYLATGAAADYFFTAWWKAIAPVFFNAFPVDRGKSKQG
ARSPRSHRGMAGSLTGDVPLLI FPEGTRSRGTGAMGTFKPGAAALAI SRGPVPIPIALVG
AWAAMPSEQAGLPKGRPSVHVAIGHPMDFVPGEIAHQFSEIRRQVIELHDQTARAYGMP
TLDEYGRHRALSQASESGDTASTNHST

A search of sequence databases reveals that the NOV67b amino acid sequence has 65 of 181 amino acid residues (35%) identical to, and 96 of 181 amino acid residues (53%) similar to, the 240 amino acid residue ptrn:TREMBLNEW-ACC:CAC01452 protein from *Streptomyces coelicolor* (PUTATIVE ACYLTRANSFERASE). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV67b is expressed in at least Bone, Bone Marrow, Brain, Liver, Lung, Lymph node, Placenta, Prostate, Thalamus, Thyroid, Uterus. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV67 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 67E.

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 15644441 ref NP_229493.1 (NC_000853)	1-acyl-sn-glycerol-3-phosphate acetyltransferase, putative [Thermotoga maritima]	247	64/214 (29%)	100/214 (45%)	3e-14
gi 9716114 emb CAC01452.1 (AL391014)	putative acyltransferase [Streptomyces coelicolor A3(2)]	240	63/184 (34%)	92/184 (49%)	e-12
gi 11693120 gb AAG38841.1 (AY010120)	putative acetyltransferase [Xanthomonas oryzae pv. oryzae]	249	68/203 (33%)	94/203 (45%)	4e-12
gi 15607028 ref NP_214410.1 (NC_000918)	2-acylglycerophosphoethanolamine acyltransferase [Aquifex aeolicus]	211	59/199 (29%)	92/199 (45%)	e-11
gi 15606303 ref NP_213682.1 (NC_000918)	long-chain-fatty-acid CoA ligase [Aquifex aeolicus]	823	47/149 (31%)	72/149 (47%)	3e-10

5 Tables 67F-G list the domain descriptions from DOMAIN analysis results against NOV67. This indicates that the NOV67 sequence has properties similar to those of other proteins known to contain this domain.

<p>Table 67F. Domain Analysis of NOV67</p> <p>gnl Pfam pfam01553, Acyltransferase, Acyltransferase. This family contains acyltransferases involved in phospholipid biosynthesis and other proteins. This family also includes tafazzin, the Barth syndrome gene.</p> <p>CD-Length = 185 residues, 95.7% aligned</p> <p>Score = 92.4 bits (228), Expect = 3e-20</p>

cigarette smoke, suggesting that cysteine modifications may have contributed to the inhibition of these two enzymes.

Although the atheroprotective role of high-density lipoprotein (HDL) has been well documented in epidemiological and animal studies, highly effective therapeutic approaches for the selective increase of plasma HDL levels or function are not yet available. Several mechanisms by which HDL exerts an atheroprotective effect have been proposed on the basis of experiments in vitro and in vivo. These mechanisms include directing excess cellular cholesterol from the peripheral tissues to the liver in 'reverse cholesterol transport', inhibiting oxidative modification or aggregation of LDL, and modulating inflammatory responses to favour vasoprotection. High density lipoproteins (HDL) mediate reverse cholesterol transport as well as the clearance of oxidation

The disclosed NOV67 nucleic acid of the invention encoding a acyltransferase-like protein includes the nucleic acid whose sequence is provided in Table 67A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 67A while still encoding a protein that maintains its acyltransferase-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 43 percent of the bases may be so changed.

The disclosed NOV67 protein of the invention includes the acyltransferase-like protein whose sequence is provided in Table 67B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table B while still encoding a protein that maintains its acyltransferase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 65 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this acyltransferase-like protein (NOV67) may function as a member of a “acyltransferase family”. Therefore, the NOV67 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV67 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the acyltransferase-like protein (NOV67) may be useful in gene therapy, and the acyltransferase-like protein (NOV67) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from Osteoporosis, Hypercalcaemia, Arthritis, Ankylosing spondylitis, Scoliosis, Hemophilia, hypercoagulation, Idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, Graft versus host, Von Hippel-Lindau (VHL) syndrome, Cirrhosis, Transplantation, Lymphedema, Allergies, Hemophilia, hypercoagulation, Idiopathic thrombocytopenic purpura, immunodeficiencies, Fertility, Osteoporosis, Hypercalcaemia, Arthritis, Ankylosing spondylitis, Scoliosis, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, or other pathologies or conditions. The NOV67 nucleic acid encoding the acyltransferase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV67 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV67 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV67 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in

[illegible]

5 A disclosed NOV68 nucleic acid of 1388 nucleotides (also referred to as CG57801-01) encoding a guanine nucleotide exchange factor-like protein is shown in Table 68A. The start and stop codons are in bold letters.

Table 68A. NOV68 nucleotide sequence (SEQ ID NO:161).

AACAGGTCACAGGAAAGTAGGAACCTGGGAGGTCTCCTGCCTCTGCTCCATGCTGCCCTC
CCTGCCCCAAGTCACCTGTCCCCTGTATGTGGGTGTCAGTTGCGAGTGAATCAGGAAGA
GCTGTGCGAAAACTCCAGCAGCACCCCCAGTGAGGAGCAGGACGAGGAGGCCAGCCAGG
CCGCGACAGACATCTGTGAGAACACAGCAGATGCGGACCAACGTATCCGGAGAGCAT
GGACACCGAGCGGGTGTATCATCAAAACCTCAGGGACATCTGTGAGGGCTATATCCGACA
GTGCCGCAAGCACACAGGAATGTTACCGTTGCGCAGCTAGCCACTATTTTGGAAACAT
TGAAGATATTTACAAATTCCAAAGAAAGTTTCTGAAAGACCTTGAGAAAACAGTACAACAA
AGAGGAAACCTCACTTAAGTGAATAGGATCTTGCTTTCTTCAAATCAAGAGGGCTTTGC
CATCTATTCCGAGTACTGCAACAACACCCGGGCGCCTGCTGGAGCTCGCCAACTCAT
GAAGCAGGGCAAGTACAGACATTTCTTTGAAGCCTGCCGCTGCTGCAGCAGATGATTGA
CATCGCCATCGACGGGTTCTTGCTCACACCAGTGCAGAAGATCTGCAAATACCCGCTGCA
GCTGGCCGAGCTGCTCAAGTATACCAACAGGAACACAGTGATTACAGCAACATAAAGGC
AGCATATGAGGCCATGAGAAATGTGGCCTGTCTGATCAACGAGCGCAAGCGCAAGCTGGA
GAGCATCGACAAGATAGCTCGCTGGCAGGTGTCTATCGTTGGGCTGGGAGGGATGGATAT
CTTAGACCGAAGCTCAGAATTGATTCATTCTGGGGAGCTGACCAAAATCACTAAGCAAGG
CAAAGCCAGCAGCGGACGTTCTTCTGTTTGACCACCAGCTGGTGTCTCTGCAGAAGGA
CTGCTGCGCAGGGACATGCTGTACTACAAGGGCCGGCTGGACATGGATGAGATGGAGCT
TGTGGACATGGGGGATTGGCGCGCAGCAAGGACTGCAACCTCAGCGTGAAAAATGCCTTCAA
GCTCGTCAGTAGGACCACAGCAGCGAGGTTTATTGTGTTTGTGCCAAAAAACAAGAAGACAA
GGCGAGGTGGCTGCAGGCCTGTGCAGATGAAAGGAGGCGGGTGCAGAGGACAAGGAGAT
GGGAATGGAATTTCAGAAAACCGAAGAAACTTGCCATGTTAAATGCTCAAAGGCAGG
ACATGGAAGTCAAAGGTAAGTTATGGAGAAGGCTTTGTCCCCTTAATGCTTATCAGTA
TTCTCCTGAAAAATGGGAGCATACCCCAAGTTGTCAGCCTGTGACCAGCTTGGAGCAAGGA
GAACAGTA

15 The disclosed NOV68 polypeptide (SEQ ID NO:162) encoded by SEQ ID NO:161 has 437 amino acid residues and is presented in Table 68B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV68 has localized in the cytoplasm with a certainty of 0.3000.

Table 68B. Encoded NOV68 protein sequence (SEQ ID NO:162).
MLPSPAPSHLS PVCGLQLRVNQEELSSENSSTPSEEQDEEASQSRHRHCENKQQMRTNVI

```
REIMDTERVYIKHLRDICEGYIRQCRKHTGMFTVAQLATIFGNIEDIYKFQRKFLKDLEK
QYNKEEPHLSEIGSCFLQNQEGFAIYSEYCNNHPGACLELANLMKQGKYRHFEEACRLLO
QMIDIAIDGFLFTPVKICKYPLQLAELLKYTTQEHSDYSNIKAAYEAMKNVACLINERK
RKLESIDKIARWQVSIVGWEGLDILDRSSELIHSGELTKITKQGKSQQRFTFFLFDHQLVS
CKKDLLRRDMLYYKGRLDMDMELVDLGDGRDKDCNLSVKNAFKLVSRTTDEVYLFCAKK
QEDKARWLQACADERRRVQEDKEMGMEISENQKKLAMLNAQKAGHGKSKGKLWRRRLCPLN
AYQYSPENGSIQVVSIL
```

A search of sequence databases reveals that the NOV68 amino acid sequence has 253 of 402 amino acid residues (62%) identical to, and 317 of 402 amino acid residues (78%) similar to, the 493 amino acid residue ptnr:SPTREMBL-ACC:Q9QX73 protein from Rattus norvegicus (Rat) (COLLYBISTIN I). Public amino acid databases include the GenBank
5 databases, SwissProt, PDB and PIR.

NOV68 is expressed in at least Kidney, Pituitary Gland, Placenta, Uterus, Aorta, Hypothalamus, Pancreas, Spleen, Epidermis, Muscle, Spinal Cord. This information was derived by determining the tissue sources of the sequences that were included in the invention
10 including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV68 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 68C.

Table 68C. BLAST results for NOV68					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 18581232 ref XP_062774.2 (XM_062774)	similar to Rho guanine nucleotide exchange factor 4, isoform a; APC-stimulated guanine nucleotide exchange factor [Homo sapiens]	652	393/399 (98%)	395/399 (98%)	0.0
gi 8809845 gb AAF79955.1 AF249745.1 (AF249745)	RhoGEF [Homo sapiens]	720	256/429 (59%)	319/429 (73%)	e-141
gi 5689561 dbj BAA83064.1 (AB029035)	KIAA1112 protein [Homo sapiens]	694	250/402 (62%)	311/402 (77%)	e-140
gi 13027402 ref NP_076447.1 (NM_023957)	collybistin I [Rattus norvegicus]	493	254/407 (62%)	318/407 (77%)	e-140
gi 7662108 ref NP_056000.1 (NM_015185)	Cdc42 guanine exchange factor 9; Cdc42 guanine exchange factor (GEF) 9 [Homo sapiens]	516	250/397 (62%)	312/397 (77%)	e-140

Table 68D. Domain Analysis of NOV68

gnl|Smart|smart00325, RhoGEF, Guanine nucleotide exchange factor for Rho/Rac/Cdc42-like GTPases; Guanine nucleotide exchange factor for Rho/Rac/Cdc42-like GTPases Also called Db1-homologous (DH) domain. It appears that PH domains invariably occur C-terminal to RhoGEF/DH domains. Improved coverage.

CD-Length = 181 residues, 100.0% aligned

Score = 156 bits (394), Expect = 3e-39

5

Table 68E. Domain Analysis of NOV68

gnl|Smart|smart00233, PH, Pleckstrin homology domain.; Domain commonly found in eukaryotic signalling proteins. The domain family possesses multiple functions including the abilities to bind inositol phosphates, and various proteins. PH domains have been found to possess inserted domains (such as in PLC gamma, syntrophins) and to be inserted within other domains. Mutations in Brutons tyrosine kinase (Btk) within its PH domain cause X-linked agammaglobulinaemia (XLA) in patients. Point mutations cluster into the positively charged end of the molecule around the predicted binding site for phosphatidylinositol lipids.

CD-Length = 104 residues, 92.3% aligned

Score = 56.2 bits (134), Expect = 4e-09

The novel protein described in this application belongs to the guanine nucleotide exchange factor family of proteins which play a significant role in signal transduction. The guanine nucleotide exchange factor (GEF) domain that regulates GTP binding protein signaling. The GEF domain regulates positively the signaling cascades that utilize GTP-binding proteins (such as those of the ras superfamily) that function as molecular switches in fundamental events such as signal transduction, cytoskeleton dynamics and intracellular trafficking. Experiments have shown that the GEF and (PH) domains of FGD1 (faciogenital dysplasia protein (FGD1)) can bind specifically to the Rho family GTPase Cdc42Hs and stimulates the GDP-GTP exchange of the isoprenylated form of Cdc42Hs. The GEF domain of FGD1 has also been shown to activate 2 kinases involved in cell proliferation; the Jun NH2-terminal kinase and the p70 S6 kinase (See Zheng et. al.; J. Biol. Chem 1996 Dec 27;271(52):33169-72). Thus this novel protein may play an important role in normal development as well as disease. This class of molecules (GEFs) is also being considered as a good drug target as the guanine nucleotide exchange factor RasGRP is a high -affinity target

for diacylglycerol and phorbol esters and is bound by bryostatin 1, a compound currently in clinical trials (See Lorenzo et. al.; Mol. Pharmacol 2000 May;57(5):840-6). Collybistin I and II, which belong to the family of dbl-like GDP/GTP exchange factors (GEFs) are most homologous to the protein described in this application. Collybistin II regulates the membrane deposition of gephyrin (an integral membrane protein) by activating a GTPase of the Rho/Rac family and may be an important determinant of inhibitory postsynaptic membrane formation and plasticity (See Kins et. al. Nat. Neurosci 2000 Jan;3(1):22-9).

The disclosed NOV68 nucleic acid of the invention encoding a guanine nucleotide exchange factor-like protein includes the nucleic acid whose sequence is provided in Table 68A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 68A while still encoding a protein that maintains its guanine nucleotide exchange factor-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 29 percent of the bases may be so changed.

The disclosed NOV68 protein of the invention includes the guanine nucleotide exchange factor-like protein whose sequence is provided in Table 68B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 68B while still encoding a protein that maintains its guanine nucleotide exchange factor-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 38 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this guanine nucleotide exchange factor-like protein (NOV68) may function as a member of a "guanine nucleotide

exchange factor family". Therefore, the NOV68 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV68 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the guanine nucleotide exchange factor-like protein (NOV68) may be useful in gene therapy, and the guanine nucleotide exchange factor-like protein (NOV68) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from cancer, trauma, regeneration (in vitro and in vivo), viral/bacterial/parasitic infections, diabetes, autoimmune disease, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, hypercalcaemia, Lesch-Nyhan syndrome, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neurodegeneration, muscular dystrophy, myasthenia gravis, atherosclerosis, aneurysm, hypertension, fibromuscular dysplasia, stroke, scleroderma, obesity, transplantation, or other pathologies or conditions. The NOV68 nucleic acid encoding the guanine nucleotide exchange factor-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV68 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV68 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV68 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in

understanding of pathology of the disease and development of new drug targets for various disorders.

NOV69

NOV69 includes two aspartate aminotransferase-like proteins disclosed below. The disclosed sequences have been named NOV69a and NOV69b.

NOV69a

A disclosed NOV60a nucleic acid of 1463 nucleotides (also referred to as CG57719-01) encoding a aspartate aminotransferase-like protein is shown in Table 69A. The start and stop codons are in bold letters.

Table 69A. NOV69a nucleotide sequence (SEQ ID NO:163).

GGAAGACTTCTGGGCAGAAGCGGAACACAGGAGCAGAGACACATAGTCTTGGCTCCAGTT
TCGTTTCAGTTATGCCACCCCTTCAGTGTTTCATGGATGTGCCCTCGCCACAAAGCTAG
AGGGCAGCTTGTTAAAGACCTACAAACAAGATGATTACCCGAACAAGATATTCTTAGCCT
ATAGAGGCACCTTCCACAGCCCCATGGAGTCCAGGAGAGATTGTTTGCAGGCTGTCTG
CAGAGCTCAGCCCTGGGGGGCCAAACCAGGCATCTGGAGCTCCCTCTGTGGTTTTCCTCA
CAGTCTGCATGACAAATGAAGGCCATCCCTGGGTTTCTCTCGTGGTGCAGAAGACTCGAC
TACAGATTTTACAGGATCCCTCCCTGAATTATGAGTACTTGCCCAACCATGGGCCTGAAAT
CATTCATCCAGGCCTCTCTAGCACTCCTCTTTGGAAAGCACAGCCAAGCCATTGTGGAGA
ACAGGGTAGGGGTGTACACACTGTGGTGACAGTGGTGCCCTTCAGCTTGGCGTCCAGT
TTCTCAGAGCTTGGCATAAGGATGCTCGTATAGTTTACATCATCTCTTCTCAAAAAGTTT
CCACAGAAGTGCATGGACTCGTCTTCCAGGACATGGGCTTTACAGTTTATGAATACTCTG
TCTGGGACCCCAAGAAGCTATGCATGGACCCGACATACTCCTCAATGTGGTGGAGCAGA
TCCCATATGGCTGTGTCTTGTGATGGGGAACATTATCGACTGCAAGTTGACACCAAGTG
GGTGGCCAAAGTTGATGTCCATGATAAAGAGCAAGCAGATATTTCCATTTTTTGTATATT
CCTGTCAAGGTTTATACACAGTGACTTGGGAAGAAGATACTAGAATCTTACAATACTTTG
TGTCTCAAGGCTTTGAGTTCTTCTGCAGCCAGTCTCTGTCCAAAATTTTGGCATTATG
ATGAAGGAGTGGGGATGCTAGTGGTGGTGGCAGTCAACAACCAGCAGCTGCTGTGTGTCC
TCTCCAGCTGGAAGGATTAGCCAGGCCCCGTGGCTAAACCCCCCAACACGGGTGCAC
GTGTCATCACCTCCATCCTCTGCAACCCTGCTCTGCTGGGAGAATGGAAGCAGAGTCTAA
AAGAAGTTGTAGAGAACATCATGCTAACCAAGGAAAAAGTGAAGGAGAACTCCAGCTCC
TGGAACCCCTGGGTCTGGGGTACATCACCGAGCAGAGTGGGACCCACGGCTATCTTG
GACTCAACTGTAAGCAGGTGGAATACCTGGTCAGGAAGAAGCACATCTATATCCCCAAGA
ACGGTCAGATTAACTTCAGCTGTATCAATGCCAACATAAATTACATCACTGAGGGCA
TCAATGAGGCTGCCTCCTCACAGAGAGCTCAGAGATGTGTCTTCCAAAGGAAAAA
CACTGATTGGAATAAACTTTAG

In a search of public sequence databases, the NOV69a nucleic acid sequence, located on chromosome 8 has 316 of 327 bases (96%) identical to a gb:GENBANK-ID:AP000501|acc:AP000501.1 mRNA from Homo sapiens (Homo sapiens genomic DNA, chromosome 8p11.2, clone:91h23 to 9-41). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV69a polypeptide (SEQ ID NO:164) encoded by SEQ ID NO:163 has 463 amino acid residues and is presented in Table 69B using the one-letter amino acid

code. Signal P, Psort and/or Hydropathy results predict that NOV69a has no signal peptide and is likely to be localized in the cytoplasm with a certainty of 0.3696.

Table 69B. Encoded NOV69a protein sequence (SEQ ID NO:164).

MPTLSVFM DVPLAHKLEGSLLKTYKDDYPNKIFLAYRGTFPQPHGVQERFVCRLS AELS
 PGGPNQASGAPSVVFLTVCM TNEGHPWVSLVVQKTRLQISQDPSLN EYLP TMGLKSF IQ
 ASLALLFGKHSQAIVENRVGGVHTVGD SGAFLGVQFLRAWHKDARIVYI ISSQKVPT EL
 HGLVFQDMGFTVYEYSVWDPKKLCMDPDILLNVVEQIPHGCVLVMGNI IDCKLT PSGWAK
 LMSMIKSKQIFPFFDIPCQGLYTS DLEEDTRILQYFVSQGF EFFCSQSLSKNFGIYDEGV
 GMLVVAVNNQQLLCVLSQLEGLAQALWLNPPNTGARVITSILCNPALLGEWKQSLKEVV
 ENIMLTKEKVKEKLQLLGTPGSWGHITEQSGTHGYLGLNCKQVEYLVRKKHIYIPKNGQI
 NFSCINANNINYITEGINEAVLLTESSEMCLPKEKKT LIGIKL

A search of sequence databases reveals that the NOV69a amino acid sequence has 163 of 228 amino acid residues (71%) identical to, and 187 of 228 amino acid residues (82%) similar to, the 264 amino acid residue ptnr:TREMBLNEW-ACC:BAB24820 protein from *Mus musculus* (Mouse) (ADULT MALE TESTIS CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY, CLONE:1700083M11, FULL INSERT SEQUENCE). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV69a is expressed in at least testis. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

NOV69b

A disclosed NOV69b nucleic acid of 1280 nucleotides (also referred to as CG57719-02) encoding a aspartate aminotransferase-like protein is shown in Table 69C. The start and stop codons are in bold letters.

Table 69C. NOV69b nucleotide sequence (SEQ ID NO:165).

CCACCCCTTTCAGTTATGCCACCCCTTTCAGTGTTTCATGGATGTGCCCCCTCGCCCAAGC
 TAGAGGGCAGCTTGTTAAAGACCTACAAACAAGATGATTACCCGAACAAGATATTCTTAG
 CCTATAGAGTCTGCATGACAAATGAAGGCCATCCCTGGGTTTCTCTCGTGGTGAGAAGA
 CTCGACTACAGATTTCACAGGATCCCTCCCTGAATTATGAGTACTTGCCACCATGGGCC
 TGAAATCATTATCCAGGCCTCTCTAGCACTCCTCTTTGGAAAGCACAGCCAAGCCATTG
 TGGAGAACAGGGCAGGGGGTGTACACACTGTTGGTGACAGTGGTGCCCTCCAGCTTGGCG
 TCCAGTTTCTCAGAGCTTGGCATAAGGATGCTCGTATAGTTTACATCATCTCTTCTCAAA
 AAGAACTGCATGGACTCGTCTTCCAGGACATGGGCTTTACAGTTTATGAATACTCTGTCT
 GGGACCCCAAGAAGCTATGCATGGACCCCGACATACTCCTCAATGTGGTGGAGCAGATCC
 CACATGGCTGTGCTCTGTGATGGGGAACATTATCGACTGCAAGTTGACACCAAGTGGGT
 GGGCAAGTTGATGTCCATGATAAAGAGCAAGCAGATATTCCCATTCTTTGATATTCCCT
 GTCAAGGTTTATACACCAAGTGAAGTGAAGATACTAGAACTTTTACAATACTTTGTGT
 CTCAAGGCTTTGAGTCTTCTGAGCCAGTCTCTGTCCAAGAAATTTTGGCATTATGATG
 AAGGAGTGGGGATGCTAGTGGTGGTGGCAGTCAACAACAGCAGCTGCTGTGTCTCTCT
 CCCAGCTGGAAGGATTAGCCAGGCCCTATGGCTAAACCCCCCAACAGGGTGCACGTG
 TCATCACCTCCATCTCTGCAACCCTGCTCTGCTGGGAGAATGGAAGCAGAGTCTAAAG
 AAGTTGTAGAGAACATCATGCTAACCAAGGAAAAAGTGAAGGAGAACTCCAGCTCCTGG
 GAACCCCTGGGTCTGGGGTACATCACCGAGCAGAGTGGGACCCACGGCTATCTTGGAC
 TCAACTCCAGCAGGTGGAATACCTGGTCAGGAAGAAGCACATCTATATCCCAAAGAAGC

GTCAGATTAACCTTCAGCTGTATCAATGCCAACATAAATTACATCACTGAGGGCATCA
ATGAGGCTGTCCTCCTCACAGAGAGCTCAGAGATGTGTCTTCCAAAGGAAAAAACAACAC
TGATTGGAATAAACTTTAG

In a search of public sequence databases, the NOV69b nucleic acid sequence, located on chromosome 8 has has 401 of 620 bases (64%) identical to a gb:GENBANK-ID:RATCASPAT|acc:D00252.1 mRNA from Rattus norvegicus (Rattus norvegicus mRNA for cytosolic aspartate aminotransferase, complete cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV69b polypeptide (SEQ ID NO:166) encoded by SEQ ID NO:165 has 421 amino acid residues and is presented in Table 69D using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV69b has no signal peptide and is likely to be localized in the cytoplasm with a certainty of 0.3645.

Table 69D. Encoded NOV69b protein sequence (SEQ ID NO:166).
MPTLSVFMVPLAHKLEGSLLKTYKQDDYPNKIFLAYRVCMTNEGHPVWSLVVQKTRLQI SQDPSLNYEYLPTMGLKSFIQASLALLFGKHSQAIVENRAGGVHTVGDGAFQLGVQFLR AWHKDARIVYI ISSQKELHGLVFDMDGFTVYEYSVWDPKKLCMDPDILLNVVEQIPHGCV LVMGNIIDCKLTPSGWAKLMSMIKSKQIFPFFDIPCGLYTSDLEEDTRILQYFVSQGF FFCSQSLSKNFGIYDEGVGMLVVVAVNNQQLLCVLSQLEGLAQALWLNPPNTGARVITS LCNPALLGEWKQSLKEVVENIMLTKEKVKEKLQLLGTGPGSWGHITEQSGTHGYLGLNSQ VEYLVRKKHIYIPKNGQINFSCINANNINYITEGINEAVLLTESSEMCLPKEKKTILIGIK L

A search of sequence databases reveals that the NOV69b amino acid sequence has have 163 of 405 amino acid residues (40%) identical to, and 236 of 405 amino acid residues (58%) similar to, the 412 amino acid residue ptrn:SWISSNEW-ACC:P17174 protein from Homo sapiens (Human) (ASPARTATE AMINOTRANSFERASE, CYTOPLASMIC (EC 2.6.1.1) (TRANSAMINASE A) (GLUTAMATE OXALOACETATE TRANSAMINASE-1)). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV69b is expressed in at least testis. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV69b polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 69E.

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 12840318 dbj BAB24820.1 (AK006984)	homolog to ASPARTATE AMINOTRANSFERASE, CYTOPLASMIC (EC 2.6.1.1) (TRANSAMINASE A) (GLUTAMATE OXALOACETATE TRANSAMINASE- 1)-putative [Mus musculus]	264	185/296 (62%)	185/296 (62%)	e-102
gi 345752 pir S29028	aspartate transaminase (EC 2.6.1.1) (clone 8C7) - human	413	224/374 (59%)	224/374 (59%)	2e-80
gi 91997 pir JT0439	aspartate transaminase (EC 2.6.1.1), cytosolic - rat	413	155/374 (41%)	224/374 (59%)	2e-80
>gi 105387 pir S13035	aspartate transaminase (EC 2.6.1.1) - human	412	155/371 (41%)	222/371 (59%)	3e-80
gi 6754034 ref NP_034454.1 (NM_010324)	glutamate oxaloacetate transaminase 1, soluble; cytosolic aspartate aminotransferase [Mus musculus]	412	154/369 (41%)	223/369 (59%)	4e-80

Table 69F-G lists the domain descriptions from DOMAIN analysis results against NOV69b. This indicates that the NOV69b sequence has properties similar to those of other proteins known to contain this domain.

5

<p>Table 69F. Domain Analysis of NOV69b</p> <p>gnl Pfam pfam00873, ACR_tran, AcrB/AcrD/AcrF family. Members of this family are integral membrane proteins. Some are involved in drug resistance.</p> <p>CD-Length = 1020 residues, 15.0% aligned Score = 51.2 bits (121), Expect = 4e-07</p>

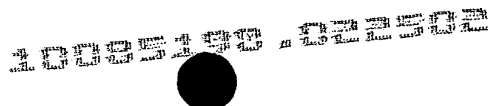


Table 69G. Domain Analysis of NOV69b

gnl|Pfam|pfam02460, Patched, Patched family. The transmembrane protein Patched is a receptor for the morphogene Sonic Hedgehog. This protein associates with the smoothed protein to transduce hedgehog signals.

CD-Length = 821 residues, 31.4% aligned

Score = 46.6 bits (109), Expect = 1e-05

Concentrations of glutamate, aspartate and glycine are significantly increased in epileptogenic cerebral cortex. The activities of the enzymes, glutamate dehydrogenase and aspartate aminotransferase, involved in glutamate and aspartate metabolism are also increased.

- 5 Polyamine synthesis is enhanced in epileptogenic cortex and may contribute to the activation of N-methyl-D-aspartate (NMDA) Receptors (See Sherwin AL (1999). *Neurochem Res* 24(11):1387-95). Nuclear magnetic resonance spectroscopy (NMRS) reveals that patients with poorly controlled complex partial seizures have a significant diminution in occipital lobe gamma aminobutyric acid (GABA) concentration. The activity of the enzyme GABA-
- 10 aminotransaminase (GABA-T) which catalyzes GABA degradation is not altered in epileptogenic cortex. NMRS studies show that vigabatrin, a GABA-T inhibitor and effective antiepileptic, significantly increases brain ABA. Glutamate decarboxylase (GAD), responsible for GABA synthesis, is diminished in interneurons in discrete regions of epileptogenic cortex and hippocampus. In vivo microdialysis performed in epilepsy surgery patients provides
- 15 measurements of extracellular amino acid levels during spontaneous seizures. Glutamate concentrations are higher in epileptic hippocampi and increase before seizure onset reaching potentially excitotoxic levels. Frontal or temporal cortical epileptogenic foci also release aspartate, glutamate and serine particularly during intense seizures or status epilepticus. GABA in contrast, exhibits a delayed and feeble rise in the epileptic hippocampus possibly
- 20 due to a reduction in the number and/or efficiency of GABA transporters. In addition; aspartate aminotransferase activity is an important index for liver function. Abnormal level and activity of aspartate aminotransferase correlates diseased liver conditions, e.g., hepatitis (See Gopal et al., (2000). *Postgrad Med* 107(2):100-2, 105-9, 113-4; Vesely et al., (1999). *Am J Med Sci* 317(6):419-24; Johnston DE (1999). *Am Fam Phys* 59(8):2223-30; Johnston SC, Pelletier LL (1997). *Medicine (Baltimore)* 76(3):185-91). Finally, aspartate aminotransferase
- 25 activity is also a marker for diagnosis of cardiovascular diseases (See Wu AH (1999). *Ann Clin Lab Sci* 29(1):18-23) and periodontal disease (See Eley BM, Cox SW (1998). *Br Dent J*

184(9):427-30). Therefore, aspartate aminotransferase is an excellent small molecule target and diagnostic marker for epilepsy, liver diseases, cardiovascular and periodontal diseases.

The disclosed NOV69 nucleic acid of the invention encoding a aspartate aminotransferase-like protein includes the nucleic acid whose sequence is provided in Table 69A or 69C or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 69A or 69C while still encoding a protein that maintains its aspartate aminotransferase-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 36 percent of the bases may be so changed.

The disclosed NOV69 protein of the invention includes the aspartate aminotransferase-like protein whose sequence is provided in Table 69B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 69B while still encoding a protein that maintains its aspartate aminotransferase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 60 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this aspartate aminotransferase-like protein (NOV69) may function as a member of a “aspartate aminotransferase family”. Therefore, the NOV69 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or

prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV69 nucleic acids and proteins of the invention are useful in potential
5 therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the aspartate aminotransferase-like protein (NOV69) may be useful in gene therapy, and the aspartate aminotransferase-like protein (NOV69) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have
10 efficacy for treatment of patients suffering from fertility, hypogonadism, or other pathologies or conditions. The NOV69 nucleic acid encoding the aspartate aminotransferase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV69 nucleic acids and polypeptides are further useful in the generation of
15 antibodies that bind immuno-specifically to the novel NOV69 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV69 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in
20 assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV70

A disclosed NOV70 nucleic acid of 4915 nucleotides (also referred to as CG-57462-
25 01) encoding a KIAA1337-like protein is shown in Table 70A. The start and stop codons are in bold letters.

Table 70A. NOV70 nucleotide sequence (SEQ ID NO: 167).	
AATATGCCTGCCATAAAGGGGAACCGGTGGTGGCAGTGGAGCTGGTGCAGGTGTGGAAAG TCATG GAGAATCCTCCTCCTCGGGTTCATCCCCAGGTTTCTCCCTTCCCCTGCCCGCGC CTGCTTGCAGGAAGCGTGTCTTACTACAGGGGCTGGGGATGCTGGCCTCACACATCCCTG CCCAGCCACAGGGTACTTCCCTGAAGCCTCCTGTACCTTCAGCCCCATCCTCGATTCTCG CCTCCGGCTCCTCCTCCCCCACGCCCTCCGGAATGAGCCCCGTACCCCCACCTCACGC GCCCTCGCTGTATAAACGCCCTCACTTGTACTGCAAGTCCCTGCGGTCCCACCTTCAGGC TTCAGCATTGCTCGACGCATCCCCAGGCCTGCTTGCTTGTCACTGTAGCGCCAGACC CAGCTTCCTTTGCGGCTCCGAGAAGCTTCCCCCTGCGACTTCCGCGAGGAGACGAGTCTG CGCAGCGTGGTGGCCGCCGCCCGCCCGACCTCTGCGCACTCTCTCCCGCGCCGGCGGCTC	

391

GGGAGATGGGGGCTGTGGAAGCCATCTCCCTGTCCATCCTCGTTGGCTCCTCCGTGGATT
 ACTGCGTCCACCTGGTCGAGGGCTACCTGCTGGCTGGAGAGAACCTGCCCCCACCAGG
 CCGAGGACGCCCCAACGCAGCGCCAGTGGCGTACGCTGGAGGCCGTGCGGCACGTGGGCG
 TGGCCATCGTCTCCAGTGCCCTCACCACGGTCATCGCCACAGTGGCCCTCTTCTTCTGCA
 TCATCGCCCCATTTGCCAAGTTCGGCAAGATTGTGGCACTCAACACGGGCGTGTCCATCC
 TCTACACGCTGACCGTCAGCACCGCCCTGCTGGGCATCATGGCGCCAGCTCTTTCACTC
 GGACCCGGACTTCCTTCCTCAAGGCCCTGGGTGCCGTGCTGCTGGCAGGGGCCCTGGGGC
 TGGGTGCCTGCCTCGTGTCTCTGCAGAGCGGCTATAAGATTCCCCTGCCCGCAGGGGCCCT
 CCCTATAGCCCCGGGACGGGCTCTGGACACTTGCACCTTTGGTCCCATGGGTGGGGGACAG
 GAGCTGCTTCCCAGCTCGACTTCAGCTAGCTGTGTCCCCAGGCCTGGGCCAGGGGCCCC
 TGCGGGCCAGCGTGGAGGCTGACACCCACACAGATGGTGTGGACCATGCTGCCTT

In a search of public sequence databases, the NOV70 nucleic acid sequence, located on chromosome 1 has 4481 of 4484 bases (99%) identical to a gb:GENBANK-ID:AB037758|acc:AB037758.1 mRNA from Homo sapiens (Homo sapiens mRNA for KIAA1337 protein, partial cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV70 polypeptide (SEQ ID NO: 168) encoded by SEQ ID NO:167 has 1561 amino acid residues and is presented in Table 70B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV70 has a signal peptide and is likely to be localized plasma membrane with a certainty of 0.8000. The most likely cleavage site for a NOV70 peptide is between amino acids 18 and 19.

Table B. Encoded NOV70 protein sequence (SEQ ID NO: 168).

MRILLGFIPRFPPLPLPAPACRKRYYRGLGMLASHIPAQPQGTSLKPPVPSAPSSILA
 SGSSSPHALRNEPRTPTLTRPCINALTCTASPCGPTFRLQHSLDASPRPACLVTVAPDP
 ASFAAPRSFPLRLPRGDESAQRGGRPPPTLCALSPAPAAQSPVRPAETMDTEDDPLLDQ
 VWLEEEQEEEEATGETFLGAQKPGPQPGAGGQCCWRHWPLASRPASGFWSTLGWAFNTP
 CCAGLVFLGCSIPMALSAFMFLYYPLDIDISYNAFEIRNHEASQRFDALTLALKSQFG
 SWGRNRDLADFTSETLQRLISEQLQLHLGNRSRQASRAPRVIIPAASLGSPGPYRDTSA
 AQKPTANRSGRLRRETPPLEDLAANQSEDPRNQRLSKNGRYQPSIPPHAAVAANQSRARR
 GASRWDYSRAYVSANTQTHAHWRIELIFLARGDAERNIFTSERLVTIHEIERKIMDHGPF
 REFCKWPHEVLKDLPLGSYSYCSPSSLMTYFFPPTERRGGKIYYDGMGQDLADIRGSLELA
 MTHPEFYWYVDEGLSADNLKSSLLRSEILFGAPLPNYYSVDDRWEQRAKFSFVVTYVA
 MLAKQSTSKVQVLYGGTDLFDYEVRRTFNNDMLLAFISSSCIAALVYILTSCSVLSFFG
 IASIGLSCLVALFLYHVVFQIYQLGILNGVAAFVIVGIGVDDVFVFINTYRQATHLEDPO
 LRMIHTVQTAGKATFFTSLTAAAYAANVFSQIPAVHDFGLFMSLIVSCCWLAVLVTMPA
 ALGLWSLYLAPLESSCQTSCHQNCRSRKTSLHFPDGFVFATPEQVGGSPAQAPIPYLDDIP
 LLEVEEEPVSLELGDVSLVSVSPEGLQPASNTGSRGHLIVQLQELLHHWVLSAVKSRWV
 IVGLFVSIILSLVFASRLRPASRAPLLFRPDNTNIQVLLDLKYNLSAEGISCITCSGLFQ
 EKPHSLQNNIRTSLEKKRRSGVWPASRPEATLQDFPGTVYISKVKSQGHPAVYRLSLNA
 SLPAWPQAVSPGDGEVPSFQVYRAPFGNFTKKLTACMSTVGLLQAASPSRKWMLTTLACD
 AKRGWKFDFFSYVATKEQQHTRKLYFAQSHKPPFHGRVCMAPPGCLSSSPDGPTKGGFF
 VPSEKVPKARLSATFGFNPCVNTGCGKPAVRPLVDTGAMVVFVFGIIGVNRTRQVDNHVI
 GDPGSVVYDSSFDLFKEIGHLCHLCKAIAANSELVKPGGAQCLPSGYSISSFLQMLHPEC
 KELPEPNLLPGQLSHGAVGVREGRVQEISMAFESTTYKGSFQTYSDYLRWESFLQQQL
 QALPEGSVLRGRFQTCHEWKQIFMEIVGVQSALCGLVLSLLICVAAVAVFTTHILLLLPV
 LLSILGIVCLVVTIMYWSGWEWGAVEAISLSILVGSSVDYCVHLVEGYLLAGENLPPHQA
 EDARTQRQWRTLEAVRHVGVAIVSSALTTVIATVPLFFCIAPFAKFGKIVALNTGVSIL
 YTLTVSTALLGIMAPSSFTRTRTSFLKALGAVLLAGALGLGACLVLLQSGYKIPLPAGAS
 L

A search of sequence databases reveals that the NOV70 amino acid sequence has 1436 of 1438 amino acid residues (99%) identical to, and 1436 of 1438 amino acid residues (99%) similar to, the 1438 amino acid residue ptnr:SPTREMBL-ACC:Q9P2K9 protein from Homo sapiens (Human) (KIAA1337 PROTEIN). Public amino acid databases include the GenBank
 5 databases, SwissProt, PDB and PIR.

NOV70 is expressed in at least Brain, Cerebral Medulla/Cerebral white matter, hippocampus, Hypothalamus, Left cerebellum, Lung, Parietal Lobe, Testis, and Right Cerebellum. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public
 10 EST sources, Literature sources, and/or RACE sources.

The disclosed NOV70 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 70C.

Table 70C. BLAST results for NOV70					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 7243055 dbj BAA92575.1 (AB037758)	KIAA1337 protein [Homo sapiens]	1438	1436/1438 (99%)	1436/1438 (99%)	0.0
gi 17448847 ref XP_052561.4 (XM_052561)	KIAA1337 protein [Homo sapiens]	1392	1388/1392 (99%)	1388/1392 (99%)	0.0
gi 5834578 emb CAB55303.1 (AL117236)	hypothetical protein [Homo sapiens]	594	592/594 (99%)	592/594 (99%)	0.0
gi 17448809 ref XP_052559.3 (XM_052559)	Similar to KIAA1337 protein [Homo sapiens]	383	361/363 (99%)	361/363 (99%)	0.0
gi 18545186 ref XP_046122.2 (XM_046122)	protein MGC13130 [Homo sapiens]	1524	54/177 (30%)	89/177 (49%)	5e-14

15 Tables 70D-E list the domain descriptions from DOMAIN analysis results against NOV70. This indicates that the NOV70 sequence has properties similar to those of other proteins known to contain this domain.

Basal cell carcinoma, medulloblastoma, rhabdomyosarcoma, and other human tumors are associated with mutations that activate the protooncogene 'Smoothed' or that inactivate the tumor suppressor 'Patched.' Smoothed and Patched mediate the cellular response to the Hedgehog secreted protein signal, and oncogenic mutations affecting these proteins cause excess activity of the Hedgehog response pathway.

Approximately 5% of patients with Gorlin syndrome develop medulloblastoma in the first few years of life, and 10% of patients with medulloblastoma diagnosed at age 2 years or under have Gorlin syndrome. Cowan et al. (1997) found that 1 out of 3 unrelated patients with medulloblastoma complicated by Gorlin syndrome had lost the wildtype allele on 9q, indicating that the Gorlin locus probably acts a tumor suppressor in the development of this tumor. They also confirmed this role in a basal cell carcinoma from the same individual. Studying patients who presented with multiple odontogenic keratocysts, Lench et al. (1997) identified 5 novel germline mutations in PTCH. Four mutations caused premature stop codons and 1 resulted in an amino acid substitution toward the C terminus of the predicted protein.

The disclosed NOV70 nucleic acid of the invention encoding a KIAA1337-like protein includes the nucleic acid whose sequence is provided in Table 70A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 70A while still encoding a protein that maintains its KIAA1337-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 1 percent of the bases may be so changed.

The disclosed NOV70 protein of the invention includes the KIAA1337-like protein whose sequence is provided in Table 70B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 70B while still encoding a protein that maintains its KIAA1337-like activities and

physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 1 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

5 The above defined information for this invention suggests that this KIAA1337-like protein (NOV70) may function as a member of a “KIAA1337 family”. Therefore, the NOV70 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein
10 therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

 The NOV70 nucleic acids and proteins of the invention are useful in potential
15 therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the KIAA1337-like protein (NOV70) may be useful in gene therapy, and the KIAA1337-like protein (NOV70) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering
20 from Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, Endocrine dysfunctions, Diabetes, obesity, Growth, Systemic lupus erythematosus, Autoimmune disease, Asthma, Emphysema,
25 Scleroderma, allergy, Fertility, ARDS, Pharyngitis, Laryngitis, Myasthenia gravis, or other pathologies or conditions. The NOV70 nucleic acid encoding the KIAA1337-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

 NOV70 nucleic acids and polypeptides are further useful in the generation of
30 antibodies that bind immuno-specifically to the novel NOV70 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV70 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in

assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV71

- 5 A disclosed NOV71 nucleic acid of 2004 nucleotides (also referred to as CG57584-01) encoding a zona pellucida glycoprotein 1 precursor-like protein is shown in Table 71A. The start and stop codons are in bold letters.

Table 71A. NOV71 nucleotide sequence (SEQ ID NO:169).
<p> CACCCACTTGCCACCTCCTCCATAAAAGGCCTGGGCCAGCTCTGGTGGCGAGGGAGTAG GGGGTGTGTCTGTGGCGTCTCATGGCAGGAGGCTCAGCCACGACCTGGGGTTACCTGTG GCCCTGCTACTGCTGGTCGCCACCCTGGGGCTGGGTAGCTACGACTGTGGGATCAAGGGA ATGCAGCTGCTGGTGTTCCTCAGGCCAGGCCAGACTCTCCCTTCAAGGTGGTGGATGAA TTTGGGAACCGATTGTGATGTCAACAACCTGCTCCATCTGCTACCACTGGGTCACCTCCAGG CCGCAGGAGCCTGCAGTCTTCTCGGCCGATTACAGAGGCTGCCACGTGCTGGAGAAGGAT GGGCGTTTCCACCTGAGGGTGTTCATGGAGGCTGTGCTGCCAATGGTCGTGTGGATGTG GCACAAGACGCTACTCTGATCTGTCCCAAACCTGACCCCTCCCGGACTCTGGACTCCAG CTGGCACCACCCGCCATGTTCTCTGTCTCAACCCACAAACCTTTCTTCTCCTCCCCACC TCTGGCCATACCTCCCAAGGCTCTGGCCATGCCTTTCCAGCCCACTGGACCCAGGGCAC AGCTCTGTCCACCCAACCCCTGCTTTACCATCCCTGGACCTGGACCTACCTTCGCCACC CTGGCTCAACCCCACTGGGGCACCTTGAACACTGGGATGTGAACAAACGAGATTACATA GGTACCCACCTGAGCCAGGAGCAGTGCCAGGTGGCCTCAGGGCACCTCCCTGCATCGTG AGAAGAACTTCAAAGAAGCCTGTGAGCAGGCTGGCTGCTGCTATGACAACACAGAGAG GTTCCCTGTTACTATGGCAACACAGCTACTGTCCAGTGCTTCAGAGATGGCTACTTCGTC CTCGTAGTGTCCCAAGAAATGGCCTTGACACACAGGATCACACTGGCCAACATCCACCTG GCCTATGCCCCCACCAGCTGCTCCCCAACACAGCACACGGAAGCTTTCGTGGTCTTCTAC TTCCCTCTCACCACCTGTGGAACCACAATGCAGGTGGCTGGCGACCAGCTCATCTATGAG AATGGCTGGTGTCTGGCATCCACATCCAAAGGGGCCACAGGGTTCCATCACGCGGGAC AGCACCTTCCAGCTTCATGTGCGCTGTGTCTTCAACGCCAGTGACTTCTGCCCATTCAG GCATCCATTTTCCACCCCCATCGCTGCTCTATGACCCAGCCCGCCCCCTGCGGCTT GAGCTGCGGATTGCCAAAGACGAGACCTGCAGCTCGTACTATGGGGAGGATGACTATCCC ATCGTGAGGGCTGTCCGAGAACCAGTCCATGTGGAGGTCCGGCTTCTGCAGAGGACAGAC CCCAACCTGGTCTGTGCTGCACCAAGTGTGGGGCGCTCCAGTGCCAACCCCTTCCAG CAGCCCCAGTGGCCCATCCTGTGACAGGATGCCCCTTCAAGGGCGACAGCTACAGAACG CAAATGGTAGCCTTGACGGGGCCACACCTTTCCAGTCGCACTACCAGCGATTCACTGTT GCTACCTTCGCCCCCTGGACTCAGGCTCCAGAGAGCCCTCAGAGGACTGGTTTACTTG TTCTGCAGCACCTCTGCCTGCCACACCTCAGGGCTGGAGACTTGCTCCACTGCATGTAGC ACTGGCACTACAAGACAGCGACGATCCTCAGGTCACCGTAATGACACTGCCAGGCCCCAG GACATCGTGAGCTCTCCGGGGCCAGTGGGCTTTGAGGATTCTTATGGGCAGGAGCCACA CTTGGGGCCACAGACTCCAATGGGAACCTCCAGCCTGAGACCTCTCCTTTGGGCGGTCTT TTGCTGCCAGCTGTTGCCCTGGTCCTTGGGTTTGGTGTCTTTGTGGGCTGAGCCAGACC TGGGCCCAGAAGCTCTGGGAAAGCAACAGACAGTGAATGGGCCCAATAACAATCATTT CAACCTACTGAAAAAAAAAAAAA </p>

- 10 In a search of public sequence databases, the NOV71 nucleic acid sequence, located on chromosome 11 has 1305 of 1704 bases (76%) identical to a gb:GENBANK-ID:MOZP1|acc:U20448.1 mRNA from Mus musculus (Mus musculus ZP1 precursor (Zp-1)

mRNA, complete cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV71 polypeptide (SEQ ID NO:170) encoded by SEQ ID NO:169 has 624 amino acid residues and is presented in Table 71B using the one-letter amino acid code.

- 5 Signal P, Psort and/or Hydropathy results predict that NOV71 has a signal peptide and is likely to be localized localized at the plasma membrane with a certainty of 0.4600. The most likely cleavage site for a NOV71 peptide is between amino acids 25 and 26.

Table 71B. Encoded NOV71 protein sequence (SEQ ID NO:170).	
MAGGSATTWGYPVALLLLVATLGLGSYDCGIKGMQLLVFPRPGQTLFPFKVVDEFGNRFDV	
NNCSICYHWVTSRPQEPVFSADYRGCHVLEKDGFRHLRVFMEAVLPNGRVDVAQDATLI	
CPKPDPSRTLDSQLAPPAMFSVSTPQTLSFLPTSGHTSQSGGHAFPSPLDPGHSSVHPTP	
ALPSPGPGPTLATLAQPHWGTLEHWDVNKRDIIGTHLSQEQCVASGHLPCIVRRTSKEA	
CQQAGCCYDNTREVPCYYGNTATVQCFRDGYFVLVVSQEMALTHRITLANIHLAYAPTSC	
SPTQHTAEAFVVFYFPLTHCGTTMQVAGDQLIYENWLVSIGIHIQKGPQGSITRDSTFQLHV	
RCVFNASDFLPIQASIFPPSPAPMTQPGPLRLELRIAKDETCSSYGEDDYPPIVRLLE	
PVHVEVRLQLQRTDPNLVLLHQCWGAPSANPFQPPQWPILSDGCPFKGDSYRTQMVALDG	
ATPFQSHYQRTTVATFALLDSGSQRALRGLVYLF CSTSACHTSGLETCSTACSTGTTRQR	
RSSGHRNDTARPQDIVSSPGPVGFEDSYGQEPTLGPTDSNGNSSLRPLLWAVLLLPAVAL	
VLGFGVFVGLSQTWAQKLWESNRQ	

- A search of sequence databases reveals that the NOV71 amino acid sequence has 422
 10 of 620 amino acid residues (68%) identical to, and 477 of 620 amino acid residues (76%) similar to, the 623 amino acid residue ptnr:SPTREMBL-ACC:Q62016 protein from Mus musculus (Mouse) (ZONA PELLUCIDA GLYCOPROTEIN 1 PRECURSOR (ZP1)). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

- NOV71 is expressed in at least Kidney, Pituitary Gland, Testis, Whole Organism.
 15 Expression information was derived from the tissue sources of the sequences that were included in the derivation of the sequence of CG57584-01. The sequence is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:MOZP1|acc:U20448.1) a closely related Mus musculus ZP1 precursor (Zp-1) mRNA, complete cds homolog in species Mus musculus :ovary. This information was
 20 derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV71 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 71C.

25

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17460872 ref XP_062198.1 (XM_062198)	similar to zona pellucida glycoprotein 1 (H. sapiens)	496	458/458 (100%)	458/458 (100%)	0.0
gi 6677653 ref NP_033606.1 (NM_009580)	zona pellucida glycoprotein 1 [Mus musculus]	623	416/628 (66%)	471/628 (74%)	0.0
gi 1113794 gb AAB60507.1 (U24230)	zona pellucida [Mus musculus]	623	414/628 (65%)	470/628 (73%)	0.0
gi 16758240 ref NP_445961.1 (NM_053509)	zona pellucida glycoprotein 1 [Rattus norvegicus]	617	416/621 (66%)	469/621 (74%)	0.0
gi 11140012 emb CAC16087.1 (AJ289697)	zona pellucida protein 1 [Gallus gallus]	934	197/369 (53%)	254/369 (68%)	2E-97

Tables 71D-E list the domain descriptions from DOMAIN analysis results against NOV71. This indicates that the NOV71 sequence has properties similar to those of other proteins known to contain this domain.

5

<p>Table 71D. Domain Analysis of NOV71</p> <p>gnl Pfam pfam00100, zona_pellucida, Zona pellucida-like domain.</p> <p>CD-Length = 266 residues, 99.6% aligned</p> <p>Score = 209 bits (533), Expect = 3e-55</p>

<p>Table 71E. Domain Analysis of NOV71</p> <p>gnl Smart smart00018, P, P or trefoil or TFF domain; Proposed role in renewal and pathology of mucous epithelia</p> <p>CD-Length = 47 residues, 100.0% aligned</p> <p>Score = 39.7 bits (91), Expect = 5e-04</p>

The mammalian zona pellucida is composed of 3 major glycoproteins, ZP1, ZP2 (182888), and ZP3 (182889). ZP3, the molecule responsible for the major sperm-receptor activity of the zona, plays a significant role in fertilization. ZP2 is implicated as a secondary sperm receptor that binds sperm only after the induction of the sperm acrosome reaction. See review by Dean (1992). The mature ZP1, ZP2, and ZP3 proteins have molecular weights of 90-110 kD, 64-76 kD, and 57-73 kD, respectively.

The disclosed NOV71 nucleic acid of the invention encoding a zona pellucida glycoprotein 1 precursor-like protein includes the nucleic acid whose sequence is provided in Table 71A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 71A while
5 still encoding a protein that maintains its zona pellucida glycoprotein 1 precursor-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or
10 complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.
15 In the mutant or variant nucleic acids, and their complements, up to about 24 percent of the bases may be so changed.

The disclosed NOV71 protein of the invention includes the zona pellucida glycoprotein 1 precursor-like protein whose sequence is provided in Table 71B. The invention also includes a mutant or variant protein any of whose residues may be changed from the
20 corresponding residue shown in Table 71B while still encoding a protein that maintains its zona pellucida glycoprotein 1 precursor-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 32 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or
25 (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this zona pellucida glycoprotein 1 precursor-like protein (NOV71) may function as a member of a “zona pellucida glycoprotein 1 precursor family”. Therefore, the NOV71 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to)
30 various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue

regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV71 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the zona pellucida glycoprotein 1 precursor-like protein (NOV71) may be useful in gene therapy, and the zona pellucida glycoprotein 1 precursor-like protein (NOV71) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from Cardiomyopathy, Atherosclerosis, Hypertension, Congenital heart defects, Aortic stenosis, Atrial septal defect (ASD), Atrioventricular (A-V) canal defect, Ductus arteriosus, Pulmonary stenosis, Subaortic stenosis, Ventricular septal defect (VSD), valve diseases, Tuberous sclerosis, Scleroderma, Obesity, Transplantation, Aneurysm, Fibromuscular dysplasia, Stroke, Anemia, Bleeding disorders, Adrenoleukodystrophy, Congenital Adrenal Hyperplasia, Diabetes, Von Hippel-Lindau (VHL) syndrome, Pancreatitis, Hyperparathyroidism, Hypoparathyroidism, SIDS, Endometriosis, Fertility, Xerostomia, Hypercalcaemia, Ulcers, Cirrhosis, Inflammatory bowel disease, Diverticular disease, Hirschsprung's disease, Crohn's Disease, Appendicitis, Hemophilia, hypercoagulation, Idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, Graft versus host, Ataxia-telangiectasia, Hemophilia, Lymphedema, Tonsillitis, Osteoporosis, Arthritis, Ankylosing spondylitis, Scoliosis, Tendinitis, Muscular dystrophy, Lesch-Nyhan syndrome, Myasthenia gravis, Dental disease and infection, Alzheimer's disease, Tuberous sclerosis, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, Growth and reproductive disorders, Endocrine dysfunctions, Systemic lupus erythematosus, Asthma, Emphysema, ARDS, Pharyngitis, Laryngitis, Hearing loss, Tinnitus, Psoriasis, Actinic keratosis, Tuberous sclerosis, Acne, Hair growth, alopecia, pigmentation disorders, cystitis, incontinence, Renal artery stenosis, Interstitial nephritis, Glomerulonephritis, Polycystic kidney disease, Systemic lupus erythematosus, Renal tubular acidosis, IgA nephropathy, Vesicoureteral reflux, glaucoma, blindness, and Hypothyroidism, or other pathologies or conditions. The NOV71 nucleic acid encoding the zona pellucida glycoprotein 1 precursor-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

antibodies that bind immuno-specifically to the novel NOV71 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV71 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV72

A disclosed NOV72 nucleic acid of 6108 nucleotides (also referred to as CG56761-01) encoding an ankyrin repeat containing protein-like protein is shown in Table 72A. The start and stop codons are in bold letters.

Table 72A. NOV72 nucleotide sequence (SEQ ID NO:171).

ATGGCCAGTCACTGGACCTGCTGACGGAGCTGCAGCTGCTGGAGAAGGTGCCACGCTG
GAGCGGCTGCGGGCTGCCAGAAGCGCGGGGCCAGCAGCTGAAGAAATGGGCACAGTAC
GAGCAGGACTTGCAGCACCAGCAAGCGAAAGCATGAGCGGAAGCGCAGCACGGGCGGCCGC
CGCAAGAAAGTGTCTTCGAGGCCAGCGTGGCCCTGCTGGAGGCCTCGCTGAGGAACGAC
GCCGAGGAAGTACGCTACTTCTTGAAGAATAAGGTACGCCCTGATTTGTGCAATGAGGAC
GGACTCACAGCCCTACACCAGTGTCTGCATCGACAACCTTTGAGGAAATTGTGAAGCTGCTC
CTCTCCCATGGTGCCAATGTGAACGCCAAGGACAACGAGCTGTGGACACCTCTCCATGCT
GCAGCCACCTGCGGCCACATCAACCTGGTGAAGATCCTCGTTTCAGTATGGGGCCGACTTG
CTTGCTGTCAACTCGGATGGGAACATGCCATATGACCTCTGCGAGGATGAACCCACCTTG
GATGTCTATCGAGACCTGCATGGCATAACAGGGCATCACCCAAGAGAAAATCAACGAGATG
CGGGTGGCTCCTGAGCAGCAGATGATTGCGGACATCCACTGCATGATCGCAGCGGGCCAG
GACCTGGACTGGATAGATGCCCAGGGTGCCACACTGCTGCACATAGCTGGAGCCATGGA
TACCTGCGGGCAGCTGAGCTCCTCCTGGATCATGGAGTGCCTGTGGATGTGAAGGACTGG
GATGGCTGGGAGCCCCTGATGCAGCTGCCTTCTGGGGACAGATGCAGATGGCAGAGCTA
TTGGTGTCCCATGGAGCTAGTCTCAGTGCAAGGACATCCATGGATGAGATGCCAATAGAC
CTGTGCGAGGAGGAAGAGTTCAAGGTCTGCTGCTGGAGCTAAAAACAAGCATGATGTG
ATCATGAAGTCACAGCTGAGGCACAAGTCATCCTTGAGCCGGAGGACCTCCAGCGCAGGC
AGCCGTGGGAAGGTGGTGCGGCGAGCCAGCCTGTGCGACAGGACCAACCTGTATAGGAAG
GAGTATGAGGGAGAGGCCATCCTGTGGCAGCGGAGTGCAGCTGAGGATCAGCGGACCTCC
ACCTACAACGGGGACATCAGGGAGACCAGGACAGACCAAGAGAATAAGGACCCCTAACCCC
AGGCTGGAGAAGCCCGTGCTACTCTCCGAATTTCTTACCAAGATCCCACGAGGTGAAGT
GACATGCCTGTTGAGAATGGCCTCCGGGCTCCGGTCACTGCCTACCAGTATGCGCTGGCC
AACGGGGATGTCTGGAAGGTGCATGAGGTGCCTGACTACAGCATGGCCTATGGCAACCCCT
GGCGTGGCCGACGCCACCCCGCCCTGGAGCAGCTACAAGGAACAGAGCCCTCAGACGCTT
CTGGAGCTGAAGCGGCAGCGGGCTGCAGCCAAGCTGCTCAGCCACCCCTTCTTAGCACA
CACCTGGGCAGCAGCATGGCCAGGACGGGCGAGAGTAGCAGTGAAGGCAAGGCCCCCTTG
ATCGGAGGCAGAACTTACCGTACAGCAGCAATGGGACCTCGGTATATTACACGGTCACC
AGCGGAGATCCCCCACTCTTAAAGTTCAAGGCCCCCATAGAGGAGATGGAGGAGAAGGTG
CATGGCTGTTGCCGTATCTCT**AGT**CTCCGTGTGATGGAGGAGGAGATGCCCTGGGGAGG
GGCTCCTGGAATCCAGGCCAGCCCAACAGCCCTGGCTGGGGAGGTGTCAGGGCAGCTGGG
GAGAGTGGGCTCTGCTTTTCAGAGGAACCTCAGACCCAGCCCTCAGCTGGCTGCCATA
GCATCCCATGTCCACGTCCTCGTGGTTCTGCTTCTGCTGCATCGTCTGCCATCTGACAC
AAGGCCTGTCTGGCCCTCCTGGTTCACTCTGCTGTCTGATCTTGGGAGGGTGGGCTTGAG

ATCCAGAGCTCTATTCTTGGTATAAAAGGCTTCTCCGGATCAGTACATGCATGTACACATTAA
CACACACACACACACATATACACACACACAAAGCTCGATCAGTGTGTGTAGGAATGA
CATACCTGGGCTCAGGGGAAGCAAGGGGGCTTAGAATTTGTGGGGTATTCCCAAAGGAT
GGAAGTTAAGACTCAGAGTCTCATACCCTGCCAATGTGGTTTTAGCAGGGGAGGGGAC
CTGCTAAGCTGAGACCCATAGTCTTCTCAGAGTTATCCCAAAGTCTGAGCCACAGCCCA
CACTGACAGGGGTGAGAAGTCTCTGCTGTGTTACAGAGGAGCCAGGAATCTACATGGGT
AGATGAGATAGACACAGACCTGCTCCCCGAGCCTTGTGTGAGAGCCACACTTCTGCCCAT
GCCAGGAGCCAGCTGTGTGACCATCCAGGGGTGGAGGGGGAAAACCAGGCAATTTCTGTTT
CTGGAATCAACCAAATCATGTTTTCTCTTGGATGGAAGTGTCAAAGGCAGAAGGGTGTG
GGAGGGGGACAAGGTCTAGTATTTACCAAAGTGTATCTGATTTTAAAAATTCTTTAGTCT
GTAAACCTCCTAGAGGGAGGGAGGTAACGAATTCATTCTTTTTGTGGATCGTATCAAG
GTCATCTGGGTTTTACTGGCTGGTGCTGGGAAAATGAAGTCAAGTGAGGAGCTTCCATTGG
AATGCTTTTTCCAGGAGAGAGGCCAGTTAATTTAAAAAAACAGTCTAGTTTAAACAGCG
ACAGAGCCCAGCAGCTGGGGTCTTTGTGAATATCCAGACTGTTTCAGCCCAGCCCATCT
CAGCCAACCCCTCCTTAGACTGAGCTGTCTCAGAGCAAGCAATTAGGGGCCAGCTGCCCTCCA
CCTCCCACCCCTTCCACCTCCATCAGTCATGTGTGCAGAGTCAGTGCTCGGGATCCCGG
GCCCAGCTTTTGCCTTTTGGGGATGCTTGGTGAGACAGATTTGCCAGTCAGCCCTTTTG
AGTTCCCGCCTCACCCAGGGGCTCCAGCCTGCACCTGCGAGGATGGTGTATGCCCCAAGT
CTGCGAATCTCAGGGTGACCTGGTCAATATCCCCCTCTGCATTCCAGGAGGCCATGGTAG
GGCTGGAGTTGGGTTCTGCCAGCCCTGCAGTTTCTAGTCCCAGCCTTCTGGTGCTGG
GGAGGGAGGACTGTGAATGGCTGTTTCTCCCTCACTGCTGAGTCTCCAGGACCCCTTT
GGAGATGCCCATGGCATGGGCACCTGCCACAGGCTCAGCCAGAACCCTCTTGGTGTACCCG
ATAAGCTGCAGGTTATCCCTTGCTCTGTGCGCCTTTTATTGTCTTAAACTACCTCCTT
AGAGCTCTGAAGGGTCTCCTAGTTCAGATTTTAAATTTGGGGAACAGATCTGGGTTCTT
TTTAAACCCTCTTCTTCTCAGTCTATGAGAACTTGCCCTGAGGGGCACCTGGGCTAGGG
GCTTGGGACTGGAAGACCATCCCCGCTTGTGCCACAACCTTTGGTCATGGGATCTGCTCT
TTGTCTATTCTTAGCCCTACTGTGGCCCCCATAGCCCATAAACCCAGAGAGGGGAGCTGG
ACTTCAGGGAGCCTGAGTGATGCTTTTCCAGGAGCAGGGCAGCTGGCTGGACCAGAAAGT
AGAGGGCCCATGGGAGTGACTGCACCCTTGGTGGCTGCTGGAAGGGGAGAGGTTCTCAGC
ATCAGGGCCACCTCCACCCAATGCCAGGATAGATGTAATCTAGAGTAGGGGTGGAGGCGG
CCCAGGAGGCTGAAGACAGGTGCACAGATGCTTCCCACGACCTTGCCATTTGGGGTGGGC
TCTTCAACATCTCAGGCTGTGGCTGGAACAGGACAGGATGATCTAAAAACACAGTACCAT
TGGCTGTAAAAAGTATGAGCCAGACTGACGCTGAAATCCCTCATGAGCCAACCTTAGC
TACAAGGTAGGGAGTTCTGAGGGGAAGCCGCGTGCTCCTCAGGAGAGAGCTGTTTAGGTTT
TCCGATCTTTTTTGCTCAGGGGCCAAACACTGAAGGCACGTAAGTCCCAACCCACTGAGCG
CCTGAGGCCATTCCCTCCTTTTCCGATGCCTCCTGCCTCCTGGGCTATTCTCTCCACC
CAGAAGGCTGGGAATCCAGCTGATTCCTTGACAGGAGCCGACTTCACACACAGGTGACT
CTCAGGCATTGGCTCATGTTTTCAGCCAGGGATAAACCATCCCTTCTTGGGGCTTTAAGT
CCCTGGGGAGCTTTCCCTGTAGGTCTCTTGGGTGTTGAGAGACAAGTTGGAGACCAACCT
CCAATGAATGAGCCGCGGTCAATTCATTAATTCACAGTAAATTTACTGAGTAGCTGCAA
CATGCCAGCCTCTACGTTAGGTTCTGCGGATAAAGGAGGAATAAGACAGAGTCAGGAGAA
CTGTTCCCTGTGGTTTCCGTCCTTGGGGACCACAGGCATCAGCAGTCCCATTCAGTCA
CCTGAGGCAAAGTGTCTGCATCTTCGTCCAGCGACCTTTGCTTTTCGGCTCCTAGAATC
CTTAGAGTCTGAATCCTTTAGCTGGGAACAGCTGTCATGGTCACCCCTGGATAACATTT
GCCACCAAGTATAGATGCTGGATCTTGGGTTCCAGGCAGACATCATCCAGGTCCATCTGG
AACTTTTCAGTGATAGTGCCTTCAGCCAGCATCTTTGGGGACTCTATAATAGCAGTTG
AGATCAGTGCTAGAGAAGCTGTTCTGCAATTTGCTGCCAAATGCATCTCAGGTTTAAAA
GTCATTGTTTTCTTGCTCATGGTGGCTCATTTATTACATAGTCCCTCACCCACTAATGG
ATAATGGGAGGAAAAGTTGCTGCTTCTTCCAGCATCAAAGCCTTCTCTTGGGAATCTGCC
TCCCTCCATGGCAGGGGTGGATTGGGAGCTGGGAGTAACCAGGCAAAGTCAACCAGATG
CCTAGCTCCTGCTGAGACCCAGGTCTATGGCAGTCTCCTCATTAGATTAAAGGAGACCAC
TTCCAAAGCAGGTGCTGCATGGCTCACCATCATATGCCCCAAACAAGTTGGCGG
TTATCACCAGACTGTGAGTTCTTGGCAAGTAGCTTGGGGAAGCTGAATAAACTCTAGGCC
CAGGGCTACTAAAGACTCTCAGGATAGAATTTCCATCAAATATACAGCATAAAGTAAACCT
GCTCTGCACTGTTTAATCCATTTCCAAGGGGCTTAGAAAAAGCTAACAAAGGGTGTGTCCCC
TGTCTGCCCCACCGGTTTGTCTGGCTTTGTAATAACATAAGACCATTGTGGTTGTGTGGTG
TCAGATACCTTCCCATCTGAGCTCTCTCACCTACCTGCTCTCTCTCCTAGAGCAGGATA
CTGGGGTACTTTTAAGAAGGGTGCTCCTTTTAAAGTGCCCAGAAAAGCTGTATTTAACTC
TTGCTATTTGTAACCTTGGGGATGGTCTCCCCTGCCCCAGGGGCACATAAGAGCAAAGGCTC
CAATGGTCACTGATGACTCTGCAAAAAGTGACCCCTGTGCCAGAAGCTATAGCCCTCTC
CCCAACAGGTCTCTTGTGTGGCCAGAGGGCCTGCTTCCCATTGGGCATTGAGGTGCCAC
CGTGCGGGGCTGGCTCTGCACACCCAGGAAAAGTCTGCAGACCCCCAGCCCTCCGCAAT

In a search of public sequence databases, the NOV72 nucleic acid sequence, located on chromosome 20 has 5593 of 5597 bases (99%) identical to a gb:GENBANK-ID:AB020630|acc:AB020630.1 mRNA from Homo sapiens (Homo sapiens mRNA for KIAA0823 protein, partial cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV72 polypeptide (SEQ ID NO:172) encoded by SEQ ID NO:171 has 567 amino acid residues and is presented in Table 72B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV72 has no signal peptide and is likely to be localized in the nucleus with a certainty of 0.9800.

MASHVDLLTELQLEKVP TLERLRAAQKRRAQQLKKWAQYEQDLQHRKRKHHERKRSTGGR
RKKVSFEASVALLEASLRNDAAEVRYFLKNKVPDL CNEDGLTALHQCCIDNFEEIVKLL
LSHGAVNNAKDNLWTP LHA AATCGHINLVKILVQYGADLLAVNSDGNMPYDLCEDEPTL
DVIETCMAYQG ITQEKINEMRVAPEQQMIADIHCMIAAGQDLWDIDAQGATLLHIAGANG
YLRAAE LLLDHGVRVDVKDWGWEPLHAAAFWGMQMAELLVSHGASLSARTSMDEMPID
LCEEEEFKVL LLELKHKHDVIMKSQLRHKSSLSRRTSSAGSRGKVVRASLSDRTNLYRK
EYEGEAILWRASAEADQSTYNGDIRETRTDQENKDPNPRLEKPVLLSEFPTKIPRGEL
DMPVENGRLRAPVSAYQYALANGDVVKVHEVPDYSMAYGNPGVADATPWSSYKEQSPQTL
LELRQRRAAKLLSHPF LSTHLGSSMARTGESSSEKAPLIGGRTPSYSSNGTSVYYTVT
SGDPPLLKFKAPIEEMEEKVHGCCRIS

A search of sequence databases reveals that the NOV72 amino acid sequence has have 412 of 412 amino acid residues (100%) identical to, and 412 of 412 amino acid residues (100%) similar to, the 412 amino acid residue ptmr:SPTREMBL-ACC:O94912 protein from Homo sapiens (Human) (KIAA0823 PROTEIN). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

404

The disclosed NOV72 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 72C.

Table 72C. BLAST results for NOV72					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 14770818 ref XP_028840.1 (XM_028840)	KIAA0823 protein [Homo sapiens]	567	567/567 (100%)	567/567 (100%)	0.0
gi 14029700 gb AAK52795.1 AF362909.1 (AF362909)	CAAX box protein TIMAP [Bos taurus]	568	550/568 (96%)	557/568 (97%)	0.0
gi 18157719 gb AAL62093.1 AF423761.1 (AF423761)	protein phosphatase 1 regulatory subunit 16B [Mus musculus]	568	547/568 (96%)	556/568 (97%)	0.0
gi 4240132 dbj BAA74846.1 (AB020630)	KIAA0823 protein [Homo sapiens]	412	412/412 (100%)	412/412 (100%)	0.0
gi 9368796 emb CAB98284.1 (AL121889)	dJ1076E17.1 (KIAA0823 protein (continues in AL023803)) [Homo sapiens]	411	411/411 (100%)	411/411 (100%)	0.0

- 5 Table 72D lists the domain descriptions from DOMAIN analysis results against NOV72. This indicates that the NOV72 sequence has properties similar to those of other proteins known to contain this domain.

Table 72D. Domain Analysis of NOV72
gnl Pfam pfam00023 , ank, Ank repeat. Ankyrin repeats generally consist of a beta, alpha, alpha, beta order of secondary structures. The repeats associate to form a higher order structure. CD-Length = 33 residues, 100.0% aligned Score = 47.4 bits (111), Expect = 2e-06

- 10 Ankyrin repeats are tandemly repeated modules of about 33 amino acids. They occur in a large number of functionally diverse proteins mainly from eukaryotes. The few known examples from prokaryotes and viruses may be the result of horizontal gene transfers. The conserved fold of the ankyrin repeat unit is known from several crystal and solution structures, e.g. from: p53-binding protein 53BP2, cyclin-dependent kinase inhibitor p19Ink4d,
- 15 transcriptional regulator GABP-beta, and NF-kappaB inhibitory protein Ikb-alpha. It has been described as an L-shaped structure consisting of a beta-hairpin and two alpha-helices. Many ankyrin repeat regions are known to function as protein-protein interaction domains.

The disclosed NOV72 nucleic acid of the invention encoding a ankyrin repeat containing protein-like protein includes the nucleic acid whose sequence is provided in Table 72A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 72A while still encoding a protein that maintains its ankyrin repeat containing protein-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 1 percent of the bases may be so changed.

The disclosed NOV72 protein of the invention includes the ankyrin repeat containing protein-like protein whose sequence is provided in Table 72B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 72B while still encoding a protein that maintains its ankyrin repeat containing protein-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 0 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this ankyrin repeat containing protein-like protein (NOV72) may function as a member of a “ankyrin repeat containing protein family”. Therefore, the NOV72 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV72 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the ankyrin repeat , containing protein-like protein (NOV72) may be useful in gene therapy, and the ankyrin repeat
5 containing protein-like protein (NOV72) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from Cardiomyopathy, Atherosclerosis, Hypertension, Congenital heart defects, Aortic stenosis, Atrial septal defect (ASD), Atrioventricular (A-V) canal defect, Ductus arteriosus , Pulmonary stenosis, Subaortic
10 stenosis, Ventricular septal defect (VSD), valve diseases, Tuberous sclerosis, Scleroderma, Obesity, Transplantation, Systemic lupus erythematosus , Autoimmune disease, Asthma, Emphysema, Scleroderma, allergy, Diabetes, Autoimmune disease, Renal artery stenosis, Interstitial nephritis, Glomerulonephritis, Polycystic kidney disease, Systemic lupus erythematosus, Renal tubular acidosis, IgA nephropathy, Hypercalcaemia, Lesch-Nyhan
15 syndrome, Von Hippel-Lindau (VHL) syndrome , Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, or other pathologies or conditions. The NOV72 nucleic acid encoding the ankyrin repeat containing protein-like protein of the invention, or fragments
20 thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV72 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV72 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods
25 known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV72 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various
30 disorders.

NOV73

A disclosed NOV73 nucleic acid of 1011 nucleotides (also referred to as CG57313-01) encoding a GPCR-like protein is shown in Table 73A. The start and stop codons are in bold letters.

5

Table 73A. NOV73 nucleotide sequence (SEQ ID NO:173).
<p>CATGGATTTCCTTAAGAAAAACAGAATGATCAATGATAGCACTTCAGTGGTTTATACTCCTTGGAT TCACAGGGCAGCCTCAGCTTCAGATGATGATCTCTGGGGTGTCTTTTCTTCTACACTATTGCCTTCAT GGGAAATATGGCCATCATCCTATTGTCTTTCCTAGATGACCATCTCCAAGTCCCATGTACTTCTTCCTT AGAAATTTGGCCATCTTGGATCTCTGTATACCACAAATATAGTCCCACAAATGTTGGTCAGTATCTGGG GCAAAGACAAAGAATTACCTTTGGTGGGTGTGCCTTTCAACTTTTCATGATGTGGCACTGTACTCAGT TGAATGCATCCTTCTGTCCATGATGTATGATCGACTCAATGCTATCTGCAAGCCTCTGCATCATATG ACCATAATGAACCTCCAACCTCTGCCAGGGCCTTGTGGTCATCTCTGGGTAGTTGGTGTGATTAATTGCA TCATACCTTCCCCTTATGCCACGAGTCTCTCTCGATGTAGGAACACCACCTAGACCACTTTTGTGTG TGTGAAATGTCTGCAAAGATCAAGATTCAAGATTGCATGTGTGGACACCACAGCCATGGAGGTAACCACA TTTGCCATGTGCGCTGATTATAGTTCTTGTTCCTCTCTTCTTATTCTTGTGTATATGTTTTCATTGCTG TGGCTGTACTCAAGATCAAGTCTGCAGCAGGAAGACAAAAGCATTGTTGGACCTGTTCTCCCATCTCGT TGTGTGATCCATCTTCTGTGGGACAGTTACATACATGTATATACAGCCAGGAAACAGTCCAAATCAGAAT GAGGGCAAACCTTCTCAGTATATTTTACTCCATTGTTACTCCAGCTTGAACCCATTAATTTATACGGTAA GGAATAAGGAGTTCAAGGGGGCCATGAAGAGGCTAACTGGAAAAGAAAAGATTGCATGGAAAAGAGG ACATTGATTCTTCTCCAGCAATTTCTAAT</p>

In a search of public sequence databases, the NOV73 nucleic acid sequence, located on chromosome 6 has has 223 of 363 bases (61%) identical to a gb:GENBANK-ID:U86270|acc:U86270.1 mRNA from Homo sapiens (Homo sapiens olfactory receptor (OR5-40) gene, partial cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV73 polypeptide (SEQ ID NO:174) encoded by SEQ ID NO:173 has 319 amino acid residues and is presented in Table 73B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV73 has a signal peptide and is likely to be localized extracellularly with a certainty of 0.6400. The most likely cleavage site for a NOV73 peptide is between amino acids 42 and 43.

Table 73B. Encoded NOV73 protein sequence (SEQ ID NO:174).
<p>MINDSHFSGFILLGFTGQPQLQMMISGVVFFFYTTAFMGNMAIILLSFLDDHLQVPMYFF LRNLAILDLCYTTNIVPQMLVSIWGDKRTIFGGCAFQLFIDVALYSVECILLSMMSYDR LNAICKPLHMTIMNLQLCQGLVVISWVGVINCIIIPSPYATSLPRCRNHLDHFFVCVK CLQSRFKIACVDTTAMEVTTTFAMCLIIIVLPLLLILVSYGFIAVAVLKIKAAGRQKAF GTCSSHLVVVSIFCGTVTYMYIQPGNSPNQNEGKLLSIFYSIVTPSLNPLIYTVRNKEFK GAMKRLTGKEKDCMEKRGH</p>

A search of sequence databases reveals that the NOV73 amino acid sequence has have 165 of 304 amino acid residues (54%) identical to, and 226 of 304 amino acid residues (74%) similar to, the 320 amino acid residue ptrn:SPTREMBL-ACC:Q9Y3N9 protein from Homo

20

sapiens (Human) (DJ88J8.1 (NOVEL 7 TRANSMEMBRANE RECEPTOR (RHODOPSIN FAMILY) (OLFACTORY RECEPTOR LIKE) PROTEIN) (HS6M1-15))). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

5 The disclosed NOV73 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 73C.

Table 73C. BLAST results for NOV73					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 18480000 gb AAL61014.1 (AY073351)	olfactory receptor MOR256-8 [Mus musculus]	308	169/307 (55%)	232/307 (75%)	8e-88
gi 13624329 ref NP112165.1 (NM_030903)	olfactory receptor, family 2, subfamily W, member 1 [Homo sapiens]	320	165/304 (54%)	226/304 (74%)	5e-86
gi 18480974 gb AAL61501.1 (AY073838)	olfactory receptor MOR256-31 [Mus musculus]	312	168/304 (55%)	227/304 (74%)	5e-86
gi 12054429 emb CAC20522.1 (AJ302602)	olfactory receptor [Mus musculus]	320	165/304 (54%)	225/304 (73%)	e-85
gi 12054431 emb CAC20523.1 (AJ302603)	olfactory receptor [Homo sapiens]	320	164/304 (53%)	226/304 (73%)	e-85

10 Table 73D lists the domain descriptions from DOMAIN analysis results against NOV73. This indicates that the NOV73 sequence has properties similar to those of other proteins known to contain this domain.

Table 73D. Domain Analysis of NOV73
gnl Pfam pfam00001 , 7tm_1, 7 transmembrane receptor (rhodopsin family).
CD-Length = 254 residues, 100.0% aligned
Score = 92.0 bits (227), Expect = 4e-20

15 The protein sequence fingerprint is potentially diagnostic of all sequences of this type in the database in which it was derived (the OWL composite sequence database, version 8.1), and has continued to perform well on subsequent database updates, identifying 240 receptors in OWL17.0. Results are compared with a commonly used pattern template for this class of receptors. The investigation suggests that discriminating power is improved in the fingerprint approach because the recognition of individual features is made mutually conditional.

Furthermore, by avoiding the definition of predetermined feature separations, members of protein families possessing all or only part of the fingerprint may be identified. PMID: 8386361

5 The fingerprint encodes the seven putative membrane-spanning motifs and was potentially diagnostic of all GPCRs (52 in all) in version 8.1 of the OWL composite sequence database, readily distinguishing them from all other integral membrane proteins. With a 3-fold increase in the size of OWL, the fingerprint has been updated and now finds 332 receptors that match all the motifs.

10 The glucagon receptor is a member of a distinct class of G protein-coupled receptors (GPCRs) sharing little amino acid sequence homology with the larger rhodopsin-like GPCR family. To identify the components of the glucagon receptor necessary for G-protein coupling, sequentially all or part of each intracellular loop (i1, i2, and i3) and the C-terminal tail of the glucagon receptor were replaced with the 11 amino acids comprising the first intracellular loop of the D4 dopamine receptor.

15 Whereas numerous mutations of the human lutropin receptor (hLHR) and human TSH receptor (hTSHR) have been shown to cause constitutive activation of these receptors, it has been suggested that either the hFSHR as a whole, or the i3/TM VI region of the hFSHR, is less susceptible to mutation-induced constitutive activation.

20 The disclosed NOV73 nucleic acid of the invention encoding a GPCR-like protein includes the nucleic acid whose sequence is provided in Table 73A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 73A while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just
25 described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least
30 in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 39 percent of the bases may be so changed.

The disclosed NOV73 protein of the invention includes the GPCR-like protein whose sequence is provided in Table 73B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 73B while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 46 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this GPCR-like protein (NOV73) may function as a member of a “GPCR family”. Therefore, the NOV73 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV73 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the GPCR-like protein (NOV73) may be useful in gene therapy, and the GPCR-like protein (NOV73) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from CNS disorders, brain disorders including epilepsy, eating disorders, schizophrenia, ADD; cancer; heart disease; inflammation and autoimmune disorders including Crohn’s disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, blood disorders; psoriasis colon cancer, leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, pancreatic disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer, or other pathologies or conditions. The NOV73 nucleic acid encoding the GPCR-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

antibodies that bind immuno-specifically to the novel NOV73 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV73 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV74

A disclosed NOV74 nucleic acid of 1008 nucleotides (also referred to as CG57315-01) encoding a GPCR-like protein is shown in Table 74A. The start and stop codons are in bold letters.

Table 74A. NOV74 nucleotide sequence (SEQ ID NO:175).

TTCACAGCTGATATTCAGAAAATAGCAAAAATGATCAATGATAGTTACTTTGGTTGGCTTATGCTCCTTG
GGTTCCCTGGGAAGCCTCAGCTGGAGATGATCATCTCTGGGGTTGTCTTTTCTCTATGCAATTTCTTT
GATGGGAAATATGGTCCTTATCCTGCTGCCATTACTGGATAAACATCTCCAAACCCCATATATTTCTTT
CTTAGAAATCTGGCTATCTTGGATCTTTGTTACACCACAAATATAGTCCCACAGATGTTGGTCAATGCCT
GGGGTAAAGACAAGAAAATCACTTTTGGTGGCTGTGCTTTTCACTTTTCACTAATGTGACGCTATGCAC
GGTTGAATGTATGCTTCTGGCTGTGATGTATATGACCCATTCAATGCTGTCTGCAAGCCTCTGGACTAT
ATGACCATAATGAACCCCAACTCTGTCAAGGCTGGTGGCCATGACCTGGTTAATTGGTGTCACTAATT
GCATGATACTTTCCCCCTGTCTGTGAGTCTTCTCGATGCGGAGACCACCACCTGGATCACTATTTTGTG
TGAAATATCTGCAATGGTCAAAATTCATGTGGGGCTACCACAGTCATGGAGTTGCATTGTGTTGTGTT
GTTGTTTTCATTTCTTGCATCACTTCTTCTCATTCTTGTGTATATGGCTTCATTGCTGTGGCTGTAC
TCAAGATCAAGTCTGCAGCAGGAAGACAAAAGCATTGGGACCTGTTTCTCCCATCTCATTGTGGTATC
CATCTTCTATGGGACTGTTAGATATATGTATATAGAGCCAGGAACAGTCCATCTCAGGATGAGGGCAAA
CTTCTCCATATATTTTACTCCATTGTACTCCACCTTGAACCCCAATCCCACTAAGGAATAAGGAGTTCA
AGTGGGCCATGAAAAGGCTTATTGGAAGAAAAGGTTCTGGAGACACAATAGGTCACTAACATCTTTT
TACAAGAAATTCCTGGCCGGGCACGGTG

In a search of public sequence databases, the NOV74 nucleic acid sequence, located on chromosome 6 has has 583 of 946 bases (61%) identical to a gb:GENBANK-ID:RATOL1RECE|acc:L34074.1 mRNA from Rattus norvegicus (Rat OL1 receptor gene, complete cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV74 polypeptide (SEQ ID NO:176) encoded by SEQ ID NO:175 has 313 amino acid residues and is presented in Table 74B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV74 has a signal peptide and is

likely to be localized extracellularly with a certainty of 0.6000. The most likely cleavage site for a NOV74 peptide is between amino acids 63 and 64.

Table 74B. Encoded NOV74 protein sequence (SEQ ID NO:176).

MINDSYFGWLMLLGFPGKPQLEMIISGVVFFFYAISLMGNMVLILLPLLDKHLQTPITYFF
LRNLAILDLCTTNIVPQMLVNAWGKDKKITFGGCAFLFTNVTLCCTVECMLLAVMSYDP
FNAVCKPLDYMTIMNPQLCQGLVAMTWLIGVTNCMILSPCPVSLPRCGDHHLDDHYFCEIS
AMVKIACGATTVMELHCVVVVVFIFLASLLILVSYGFIHAVVLKIKSAAGRQKAFGTCTF
SHLIVVSIFYGTVMYIEPGNSPSQDEGKLLHIFYSIVTPTLNPIPLRNKEFKWAMKRL
IGKEKSGSDTIGH

A search of sequence databases reveals that the NOV74 amino acid sequence has 161 of 298 amino acid residues (54%) identical to, and 220 of 298 amino acid residues (73%) similar to, the 320 amino acid residue ptnr:SPTREMBL-ACC:Q9Y3N9 protein from Homo sapiens (Human) (DJ88J8.1 (NOVEL 7 TRANSMEMBRANE RECEPTOR (RHODOPSIN FAMILY) (OLFACTORY RECEPTOR LIKE) PROTEIN) (HS6M1-15))). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV74 is expressed in at least Whole Organism. Expression information was derived from the tissue sources of the sequences that were included in the derivation of the sequence of CG57315_01. The sequence is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:RATOL1RECE|acc: L34074.1) a closely related Rat OL1 receptor gene, complete cds homolog in species Rattus norvegicus: heart. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV74 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 74C.

Table 74C. BLAST results for NOV74

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17464347 ref XP_069460.1 (XM_069460)	similar to dM538M10.7 (novel 7 transmembrane receptor (rhodopsin family) (olfactory receptor like) protein) [Homo sapiens]	590	186/186 (100%)	186/186 (100%)	e-100

<u>gi 18480974 gb AAL61501.1 </u> (AY073838)	olfactory receptor MOR256-31 [Mus musculus]	312	167/293 (56%)	218/293 (73%)	e-84
<u>gi 18480000 gb AAL61014.1 </u> (AY073351)	olfactory receptor MOR256-8 [Mus musculus]	308	159/303 (52%)	225/303 (73%)	2e-83
<u>gi 13624329 ref NP112165.1 </u> (NM_030903)	olfactory receptor, family 2, subfamily W, member 1 [Homo sapiens]	320	161/301 (53%)	219/301 (72%)	3e-83
<u>gi 12054429 emb CAC20522.1 </u> (AJ302602)	olfactory receptor [Homo sapiens]	320	161/301 (53%)	218/301 (71%)	6e-83

Table 74D lists the domain descriptions from DOMAIN analysis results against NOV74. This indicates that the NOV74 sequence has properties similar to those of other proteins known to contain this domain.

5

<p align="center">Table 74D. Domain Analysis of NOV74</p> <p align="center"><u>gnl Pfam pfam00001</u>, 7tm_1, 7 transmembrane receptor (rhodopsin family).</p> <p align="center">CD-Length = 254 residues, 99.2% aligned</p> <p align="center">Score = 96.3 bits (238), Expect = 2e-21</p>

The disclosed NOV74 nucleic acid of the invention encoding a GPCR-like protein includes the nucleic acid whose sequence is provided in Table 74A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 74A while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 39 percent of the bases may be so changed.

The disclosed NOV74 protein of the invention includes the GPCR-like protein whose sequence is provided in Table 74B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 74B while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 46 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this GPCR-like protein (NOV74) may function as a member of a "GPCR family". Therefore, the NOV74 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV74 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the GPCR-like protein (NOV74) may be useful in gene therapy, and the GPCR-like protein (NOV74) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from CNS disorders, brain disorders including epilepsy, eating disorders, schizophrenia, ADD; cancer; heart disease; inflammation and autoimmune disorders including Crohn's disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, blood disorders; psoriasis colon cancer, leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, pancreatic disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer, or other pathologies or conditions. The NOV74 nucleic acid encoding the GPCR-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

antibodies that bind immuno-specifically to the novel NOV74 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-
5 NOVX Antibodies” section below. The disclosed NOV74 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

10 NOV75

A disclosed NOV75 nucleic acid of 1050 nucleotides (also referred to as CG57317-01) encoding a GPCR-like protein is shown in Table 75A. The start and stop codons are in bold letters.

Table 75A. NOV75 nucleotide sequence (SEQ ID NO:177).

```
ACTTTTAAAGATTGATATTTTGCCCAATGGCCAACACATTATCCTCCCTGAATTCTTGTAATGTGTTTC
TCCTAGTTCTGAACAGGGTGATGGGCATGACCAACAGCAGTGTCAAGGGAGACTTCATCCTGGTGGGTTT
CTCTCATCAGCCCCACCTGGAAAAGATCCTCTTTGTGGCTGTTTTGATATCCTATCTCCTTACCCTTG
GGAAATACAGTAATTATTCTGATCTGCTCTGTAGACCCTAAACTCAAGACACCCATGTATTTTTTCTTAA
CTCACCTCTCCTTAGTTGATATCTGTTTTACCACCAGTATTGTCCCCAGCTGCTGTGGAACCTAAAAGG
ACCTGACAAAACAATCACATTCCTGGGTTGTGTCTATCCAGCTCTACATCTCCCTGGCATTGGGCTCCACT
GAGTGTGTCCTCCTGGCTGTAATGGCTTTTGATCGCTATGCTGCAGTTTGCAAACCTCTCCACTATACCG
CCGTAATGAACCCCTCAGCTGTGCCAGGCTCTGGCAGGGGTTGCGTGGCTGAGTGGAGTGGGAAACACTCT
TATCCAGGGCACTGTACCCCTCTGGCTTCCTCGCTGTGGACACCGATTGCTCCAACATTTCTTCGTGAG
GTACCCCTCCATGATTAAGCTTGCAATGTGTGGACATCCATGATAATGAGGTTGAGTCTTTGTTGCTTCAC
TGGTCTTGCTCCTCTTGCCCTTAGTGCTAATACTGCTGTCTATGGACATATAGCCAGGTGGTCATAAG
GATCAAGTCAGTCCAGGCTGGTGCAAAGGCTGGGGACATGTGGATCCCATTTGATAGTAGTGTCCCTC
TTCTGTGGGACCATCACAGCTGTCTACATCCAGTCCAACAGTTCTTATGCCCATGCTCATGGGAAGTTCA
TCTCCCTCTTCTATACAGTTGTGACCCGACCCTCAATCCTCTCATCTACACACTGAGGAATAATGACGT
GAAAGGAGCACTGCGATTATTTAACAGAGACTTAGGCACATAAAAAATGAAGCAGAGTACACAGCGCTCA
```

In a search of public sequence databases, the NOV75 nucleic acid sequence, located on chromosome 6 has 611 of 912 bases (66%) identical to a gb:GENBANK-
ID:HUMORLMHC|acc:L35475.1 mRNA from Homo sapiens (Human olfactory receptor-like
gene, complete cds). Public nucleotide databases include all GenBank databases and the
20 GeneSeq patent database.

The disclosed NOV75 polypeptide (SEQ ID NO:178) encoded by SEQ ID NO:177 has 331 amino acid residues and is presented in Table 75B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV75 has a signal peptide and is

likely to be localized extracellularly with a certainty of 0.6400. The most likely cleavage site for a NOV75 peptide is between amino acids 23 and 24.

Table 75B. Encoded NOV75 protein sequence (SEQ ID NO:178).

MANTLSSLNSCNVFLVLNLRVGMGTNSSVKGDFILVGFSSHQPHLEKILFVAVLISYLLTL
VGNTVILILICSVDPKLKTPMYFFLTHLSLVDICFTTSIVPQLLWNLKGPDKTITFLGCVI
QLYISLALGSTECVLLAVMAFDHYAAVCKPLHYTAVMNPQLCQALAGVAWLSGVGNTLIQ
GTVTLWLPRCGHRLQLHFLREVPSMIKLACVDIHDNEVQLFVASLVLLLLPLVLILLSYG
HIAKVIRIKSVQAWCKGLGTCGSHLIVVSLFCGTITAVYIQSNSSYAHAGKFIISLFYT
VVTPTLNPLIYTLRNNDVKGALRLFNRLDGT

A search of sequence databases reveals that the NOV75 amino acid sequence has 178 of 306 amino acid residues (58%) identical to, and 234 of 306 amino acid residues (76%) similar to, the 312 amino acid residue ptrn:SPTREMBL-ACC:Q9R0Z2 protein from Mus musculus (Mouse) (573K1.3 (MM17M1-4 (NOVEL 7 TRANSMEMBRANE RECEPTOR (RHODOPSIN FAMILY) (OLFACTORY RECEPTOR LIKE) PROTEIN))). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

The disclosed NOV75 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 75C.

Table 75C. BLAST results for NOV75

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17464347 ref XP_069460.1 (XM_069460)	similar to dM538M10.7 (novel 7 transmembrane receptor (rhodopsin family) (olfactory receptor like) protein) [Homo sapiens]	590	270/322 (83%)	271/322 (83%)	e-121
gi 18480232 gb AAL61130.1 (AY073467)	olfactory receptor MOR256-9 [Mus musculus]	309	254/308 (82%)	273/308 (88%)	e-120
gi 18480518 gb AAL61273.1 (AY073610)	olfactory receptor MOR256-19 [Mus musculus]	317	229/298 (76%)	259/298 (86%)	e-112
gi 18565094 ref XP_094939.1 (XM_094939)	hypothetical protein XP_094939 [Homo sapiens]	310	225/235 (95%)	226/235 (95%)	e-112
gi 14596252 emb CAC43450.1 (AL136158)	dM538M10.7 (novel 7 transmembrane receptor (rhodopsin family) (olfactory receptor like) protein) [Mus musculus]	317	233/307 (75%)	260/307 (83%)	e-111

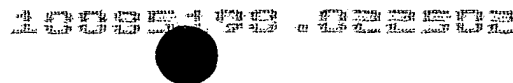


Table 75D lists the domain descriptions from DOMAIN analysis results against NOV75. This indicates that the NOV75 sequence has properties similar to those of other proteins known to contain this domain.

Table 75D. Domain Analysis of NOV75	
<u>gnl Pfam pfam00001</u> , 7tm_1, 7 transmembrane receptor (rhodopsin family).	
CD-Length = 254 residues, 46.9% aligned	
Score = 78.6 bits (192), Expect = 5e-16	

5

The disclosed NOV75 nucleic acid of the invention encoding a GPCR-like protein includes the nucleic acid whose sequence is provided in Table 75A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 75A while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 34 percent of the bases may be so changed.

10

15

20

The disclosed NOV75 protein of the invention includes the GPCR-like protein whose sequence is provided in Table 75B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 75B while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 42 percent of the residues may be so changed.

25

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this GPCR-like protein (NOV75) may function as a member of a “GPCR family”. Therefore, the NOV75 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV75 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the GPCR-like protein (NOV75) may be useful in gene therapy, and the GPCR-like protein (NOV75) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from CNS disorders, brain disorders including epilepsy, eating disorders, schizophrenia, ADD; cancer; heart disease; inflammation and autoimmune disorders including Crohn’s disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, blood disorders; psoriasis colon cancer, leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, pancreatic disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer, or other pathologies or conditions. The NOV75 nucleic acid encoding the GPCR-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV75 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV75 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV75 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV76

A disclosed NOV76 nucleic acid of 1063 nucleotides (also referred to as CG57321-01) encoding a GPCR-like protein is shown in Table 76A. The start and stop codons are in bold letters.

5

Table 76A. NOV76 nucleotide sequence (SEQ ID NO:179).

TCTCTAATTCTCAGTGGCTTCCTCCTACTGTTGATGTCTATCCCTAACTGTGGGTATTTAGAGGTCTCAG
CTGGAATTTTACCTCCAGTGCTAACATGTGGATCAACAATCAAAGCTCGCTAGATGATTTTATCCTATT
GGGATTTTCTGACCGTCCCTGGCTAGAGACACCCCTCTGTGTAATCTTTCTGGTGGCCTACATCTTTTCC
CTATTTGGAAATATCTCCATTATCCTAGTTTCCCATCTGGATCCCAGCTTGACAGTCCCATGTACTTTT
TTGTCTCTAATCTATCCTTTCTGGACCTCTGCTATACCACCAGCACTGTCCACAGATGCTGGTCAACCT
CCGGGGACCAGAAAAGACCATTAGCTATGGGGGTTGTGTGCCCAACTCTATATATTTTGGCCCTGGGT
TCTACTGAATGCATACTTCTAGCCATCATGGCCTTTGACCGTTACGCTGCCATATGCAAGCCCTTCACT
ACCCAGTCATCATGAACCATAGACGCTGTATCCACATGGCTGCTGGCACTTGGATCAGTGGCTTTGCTAA
CTCCCTTGTCCAGTCCACTCTCACAGTGGTGGCCCCAAGATGTGGACAGAGGGTGTGGACCATTCTTCT
TGTGAAGTTCAGCCCTTTTGAAGTACGCTGTATTGATATTCGTGTGAATGAAATGGAGCTCAATGTAC
TAGGCGCTTTGCTTCTCCTGATGCCACTCACCCCTCATCTGGGCACTTATGTGTTCAATTGCTCAGGCAGT
AATGAGAATCTGCTCTGCTGAAAGTCGCTGGAAGGCTTCAATACCTGTGCCTCACATTGCTGGTGGTC
TCCCTCTTCTACTTCACAGCCATCAGTATGTATGTCCAGCCTCCCTCTAGCTATTCTCATGACCGGGGGA
AGATCATCATGGCTCTCTTTTATGGCATTGTACACCCACCCTCAACCCATTCTACACATTGAGAAA
CAAGGATGTGAAAGCTGCCCTGAGAAGGTCACCTGACTAAAGAGTTTGGATTAAGACAAGATGATATCTG
AAAAGAAGTCCTA

In a search of public sequence databases, the NOV76 nucleic acid sequence, located on chromosome 6 has 657 of 971 bases (67%) identical to a gb:GENBANK-ID:RATOL1RECE|acc:L34074.1 mRNA from Rattus norvegicus (Rat OL1 receptor gene, complete cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

10

The disclosed NOV76 polypeptide (SEQ ID NO:180) encoded by SEQ ID NO:179 has 336 amino acid residues and is presented in Table 76B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV76 has a signal peptide and is likely to be localized localized extracellularly with a certainty of 0.6000. The most likely cleavage site for a NOV76 peptide is between amino acids 14 and 15.

15

Table 76B. Encoded NOV76 protein sequence (SEQ ID NO:180).

MSIPNCGYLEVSAGISPPSANMWINNQSSLDFFILLGFSDRPWLETPLCVIFLVAYIFSL
FGNISIIILVSHLDPQLDSPMYFFVSNLSFLDLCYTTSTVPQMLVNLRGPEKTISYGGCVA
QLYIFLALGSTECILLAIMAFDRYAAICKPLHYPVIMNHRRCIHMAAGTWISGFANSLVQ
STLTVVAPRCGQVRVLDHFFCEVPALLKLACIDIRVNEMELNVLGALLLMLPLTLILGTYV
FIAQAVMRICSAESRWKAFNTCASHLLVVSIFYFTAISMYVQPPSSYSHDRGKIIMALFY
GIVTPTLNPFIYTLRNKDVKAALRRSLTKEFWIKTR

A search of sequence databases reveals that the NOV76 amino acid sequence has 197 of 308 amino acid residues (63%) identical to, and 245 of 308 amino acid residues (79%)

similar to, the 313 amino acid residue ptmr:SPTREMBL-ACC:Q63394 protein from Rattus norvegicus (Rat) (OL1 RECEPTOR). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

5 The disclosed NOV76 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 76C.

Table 76C. BLAST results for NOV76					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17464349 ref XP_069461.1 (XM_069461)	similar to olfactory receptor, family 2, subfamily B, member 2 [Homo sapiens]	298	297/315 (94%)	298/315 (94%)	e-147
gi 18479350 gb AAL60689.1 (AY073026)	olfactory receptor MOR256-3 [Mus musculus]	315	287/315 (91%)	302/315 (95%)	e-147
gi 11177906 ref NP_068632.1 (NM_021860)	olfactory receptor [Mus musculus]	313	197/308 (63%)	245/308 (78%)	e-104
gi 14780900 ref NP_149046.1 (NM_033057)	olfactory receptor, family 2, subfamily B, member 2 [Homo sapiens]	357	199/307 (64%)	245/307 (78%)	e-103
gi 18480406 gb AAL61217.1 (AY073554)	olfactory receptor MOR256- 10 [Mus musculus]	313	196/308 (63%)	242/308 (77%)	e-103

10 Table 76D lists the domain descriptions from DOMAIN analysis results against NOV76. This indicates that the NOV76 sequence has properties similar to those of other proteins known to contain this domain.

Table 76D. Domain Analysis of NOV 76
gnl Pfam pfam00001 , 7tm_1, 7 transmembrane receptor (rhodopsin family).
CD-Length = 254 residues, 100.0% aligned
Score = 95.9 bits (237), Expect = 3e-21

15 The disclosed NOV76 nucleic acid of the invention encoding a GPCR-like protein includes the nucleic acid whose sequence is provided in Table 76A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 76A while still encoding a protein that maintains

its GPCR-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

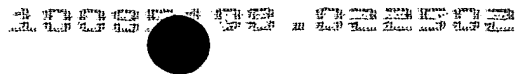
In the mutant or variant nucleic acids, and their complements, up to about 33 percent of the bases may be so changed.

The disclosed NOV76 protein of the invention includes the GPCR-like protein whose sequence is provided in Table 76B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 76B while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 37 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this GPCR-like protein (NOV76) may function as a member of a "GPCR family". Therefore, the NOV76 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV76 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the GPCR-like protein (NOV76) may be useful in gene therapy, and the GPCR-like protein (NOV76) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering



from CNS disorders, brain disorders including epilepsy, eating disorders, schizophrenia, ADD; cancer; heart disease; inflammation and autoimmune disorders including Crohn's disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, blood disorders; psoriasis colon cancer, leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, pancreatic disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer, or other pathologies or conditions. The NOV76 nucleic acid encoding the GPCR-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV76 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV76 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV76 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV77

A disclosed NOV77 nucleic acid of 1014 nucleotides (also referred to as CG57419-01) encoding a GPCR-like protein is shown in Table 77A. The start and stop codons are in bold letters.

Table 77A. NOV77 nucleotide sequence (SEQ ID NO:181).

CATCTTTAGTGTGGTCCTTGGAAAGCTCATGGGTCAGGAAAATAAAAACCAGACATGGGTGAGTGAGTTCA TTCTGCTGGGGATTTCCAGTGATGGGGCATTGAGGTATCCCTCTTCGCCCTGATCCTGGCCATGTATTT GGTGACTATTTAGGAAACACCCTCATTCTTCTTCTGATCAGACTGGACAACAGGCTTCATACCCCCATG TACTTCTCCCTTAGTGTTCTGTGTCATTTGTGGACTTTTGTATACAAAGAGTATGTGCCACAAATGCTGT CCCACTTGCTCTCAGCCCGAAAGTCCATCCCATTCTACAGTTGTGTGCTCCAGCTCTATGTTTCTCTGGC ATTGTGTGGGTCTGAGTTCTTCTGCTGGGGGCCATGGCCTATGACCGCTACGTGGCCGTGTGCCACCCA CTGCACTACACGGTCATCATGATGGAGGGCTGTGCCTGGGGCTGGCGGCCAGCCGCTGGTGGCTGGCT TCTCAAATCCCTGATGGAACAATTATCACCTTCCAGCTTTTATCACCTTCCAGCTTCCCTGTGTACGG TGTTATCAATCACTTTGTCTGTGAGACCTTAGCAGTGCTACAGCTAGCCTGTGTGGATGTCCCTTCAAC AAGGTCATGGTGGCCATCTCAGGGTTTCTGGTGATCTTGCTTCCCTGTTCCCTGGTTCTATTCTCCTATG CTGTCATAGTTGCCACCATTTGTGCATTGTTCTACCCAGGTACGCTGCAAAGCCTTTGGGACCTGTGC CTCTCACCTCATTTGTGGTTTGCATGTGCTTTGGGGCTACCATGTCACCTACCTGGGGCCACAGTTGGCC TCCTCAGCAGAGGAAGAGAAGATGATTGCTCTCTTCTATGGAGTGGTGTACCCCATGTTGAACCCCTTGA TCTACAGCTTGAGGAATAAGGAAGTTACGGCTGCTGTCCGGAAGTTTGTAGAAAGATGCAGATAAAGGGT CAAGACTCTAAGAACCTCTTGTTATCTATCATCA

In a search of public sequence databases, the NOV77 nucleic acid sequence, located on chromosome 7 has 358 of 495 bases (72%) identical to a gb:GENBANK-ID:HSU56421|acc:U56421.1 mRNA from Homo sapiens (Human olfactory receptor (OLF3) gene, complete cds). Public nucleotide databases include all GenBank databases and the
5 GeneSeq patent database.

The disclosed NOV77 polypeptide (SEQ ID NO:182) encoded by SEQ ID NO:181 has 315 amino acid residues and is presented in Table 77B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV77 has a signal peptide and is likely to be localized extracellularly with a certainty of 0.6000. The most likely cleavage site
10 for a NOV77 peptide is between amino acids 43 and 44.

Table 77B. Encoded NOV77 protein sequence (SEQ ID NO:182).	
MGQENKNQTWVSEFILLGISSDWGIQVSLFALILAMYLVTILGNTLILLIRLDNRLHTP MYFSLSVLSFVDFCYTKSIVPQMLSHLLSARKSIPFYSCVLQLYVSLALCGSEFFLLGAM AYDRYVAVCHPLHYTVIMHGGLCLGLAASRLVAGFSNSLMETIITFQLLSPSSFLCHGVI NHFVCETLAVLQLACVDVFPFNKVMVAISGFLVILLPCSLVLFYSYACIVATILCIRSTQVR CKAFGTCASHLIVVCMCFGATICTYLGPQLASSAEEEEKMIALFYGVVSPMLNPLIYSLRN KEVTA AVRKVLRCR	

A search of sequence databases reveals that the NOV77 amino acid sequence has have 180 of 305 amino acid residues (59%) identical to, and 229 of 305 amino acid residues (75%)
15 similar to, the 317 amino acid residue ptrn:SWISSPROT-ACC:Q95156 protein from Canis familiaris (Dog) (OLFACTORY RECEPTOR-LIKE PROTEIN OLF3). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

The disclosed NOV77 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 77C .

20

Table 77C. BLAST results for NOV77					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 18480604 gb AAL61316.1 (AY073653)	olfactory receptor MOR257-3 [Mus musculus]	310	239/315 (75%)	268/315 (84%)	e-112
gi 2495055 sp Q95156 OLF3 CANFA	OLFACTORY RECEPTOR-LIKE PROTEIN OLF3	317	178/305 (58%)	226/305 (73%)	E-82

<u>gi 6912558 ref NP_036501.1 </u> (NM_012369)	olfactory receptor, family 2, subfamily F, member 1; olfactory receptor, family 2, subfamily F, member 5; olfactory receptor, family 2, subfamily F, member 4 [Homo sapiens]	317	173/305 (56%)	221/305 (71%)	e-79
<u>gi 9297120 sp Q13607 O2F1 HUMAN</u>	OLFACTORY RECEPTOR 2F1 (OLFACTORY RECEPTOR-LIKE PROTEIN OLF3)	317	173/305 (56%)	221/305 (71%)	e-79
<u>gi 18479500 gb AAL60764.1 </u> (AY073101)	olfactory receptor MOR257-1 [Mus musculus]	313	176/305 (57%)	219/305 (71%)	3e-79

Table 77D lists the domain descriptions from DOMAIN analysis results against NOV77. This indicates that the NOV77 sequence has properties similar to those of other proteins known to contain this domain.

5

<p align="center">Table 77D. Domain Analysis of NOV77</p> <p align="center"><u>gnl Pfam pfam00001</u>, 7tm_1, 7 transmembrane receptor (rhodopsin family).</p> <p align="center">CD-Length = 254 residues, 94.1% aligned Score = 75.5 bits (184), Expect = 4e-15</p>

The disclosed NOV77 nucleic acid of the invention encoding a GPCR-like protein includes the nucleic acid whose sequence is provided in Table 77A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 77A while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be

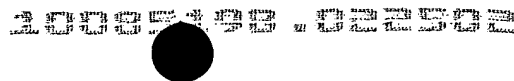
used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 28 percent of the bases may be so changed.

The disclosed NOV77 protein of the invention includes the GPCR-like protein whose sequence is provided in Table 77B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 77B while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 41 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this GPCR-like protein (NOV77) may function as a member of a "GPCR family". Therefore, the NOV77 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV77 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the GPCR-like protein (NOV77) may be useful in gene therapy, and the GPCR-like protein (NOV77) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from CNS disorders, brain disorders including epilepsy, eating disorders, schizophrenia, ADD; cancer; heart disease; inflammation and autoimmune disorders including Crohn's disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, blood disorders; psoriasis colon cancer, leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, pancreatic disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer, or other pathologies or conditions. The NOV77 nucleic acid encoding the GPCR-like protein of the invention, or fragments thereof, may further be useful in diagnostic



applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV77 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV77 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV77 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV78

A disclosed NOV78 nucleic acid of 1151 nucleotides (also referred to as CG57425-01) encoding a GPCR-like protein is shown in Table 78A. The start and stop codons are in bold letters.

Table 78A. NOV78 nucleotide sequence (SEQ ID NO:183).

GGACACTGGTTTGGGCCATATGGATGGTGAGCAATGTATGACCTGATTCTGTTTCATTAAG
ATAAGCTTTATGTCTCCTACTCTAAGAACTCTTCAATTTTCATCATTCTCACCTTTTGC
CTTTAGGTTTTCCGAAGGTCAACA**ATG**AAAAACAGAACCATGTTTGGTGAGTTTATTCTA
CTGGGCCTTACAAATCAACCTGAAC**TCCA**AGTGATGATATTTCATCTTTCTGTTCCCTCACC
TACATGCTAAGTGCTCTAGGAAATCTGACTATTATCACCCCTCACCTTACTAGACCCCCAC
CTCCAGACCCCCATGTATTTCTTCCCTCCGGAATTTCTCCTTCTTAGAAATTTCTTCACA
TCCATTTTTATTTCCAGATTTCTGACCAGCATGACAACAGGAAATAAAGTTATCAGCTTT
GCTGGCTGCTTGACTCAGTATTTTTTTGCTATATTTCTTGGAGCTACCGAGTTTTACCTC
CTGGCCTCCATGTCTTATGATCGTTATGTGGCCATCTGCAAACCCCTTGCATTACCTGACT
ATTATGAGCAGCAGAGTCTGCATACAAC**TAGT**GTTCGCTCCTGGTTGGGGGGATTCCTA
GCAATCTTACCACCAATCATCCTGATGACCCAGGTAGATTTCTGTGTCTCCAACATTCTG
AATCACTATTACTGTGACTATGGGCCTCTCGTGGAGCTTGCCTGCTCAGACACAAGCCTC
TTAGAACTGATGGTCATCCTCTTGGCCGTGTGTGACTCTCATGGTTACTCTGGTGCTGGTG
ACACTTCTTACACATACATTATCAGGACTATTCTGAGGATCCCTTCCGCCAGCAAAGG
ACAAAGGCCTTTTCCACTTGTTCCTCCACATGATTGTTCATCTCCCTCTCTTATGGCAGC
TGCATGTTTATGTACATTAATCCTTCTGCAAAAGAAGGAGGTGCTTTCAACAAAGGAATA
GCTGTACTCATTACTTCGGTTACTCCCTTACTGAATCCCTTCATATATACTTTAAGAAAT
CAGCAAGTGAAACAAGCTTTCAAGGACTCAGTCAAAAGATTGTGAAACTTT**TAAAAA**AGG
AGATTACACTTCAAAATACATTTTCACTTAACAAATATGCATTGAATGTCTATATTTCAA
GTGCTAAATTG

In a search of public sequence databases, the NOV78 nucleic acid sequence, located on chromosome 12 has 605 of 931 bases (64%) identical to a gb:GENBANK-

ID:AF102523|acc:AF102523.1 mRNA from Mus musculus (Mus musculus olfactory receptor

C6 gene, complete cds). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV78 polypeptide (SEQ ID NO:184) encoded by SEQ ID NO:183 has 309 amino acid residues and is presented in Table 78B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV78 has a signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.6000. The most likely cleavage site for a NOV78 peptide is at amino acid 39.

Table 78B. Encoded NOV78 protein sequence (SEQ ID NO:184).	
MKNRTMFGEFILLGLTNQPELQVMIFIFLFLTYMLSVLGNLTIITLTLDDPHLQTPMYFF	
LRNFSFLEISFTSIFIPRFLTSMTTGNKVISFAGCLTQYFFAIFLGATEFYLLASMSYDR	
YVAICKPLHYLTIMSSRVCIQLVFCSWLGGFLAILPPIILMTQVDFCVSNILNHYYCDYG	
PLVELACSDTSLLELMVILLAVVTLMVTLVLVLTLSYTYIIRTLRIPSAQQRTKAFSTCS	
SHMIVISLSYGSCFMFYINPSAKEGGAFNKGIAVLITSVTPLLNPFIYTLRNQQVKQAFK	
DSVKKIVKL	

A search of sequence databases reveals that the NOV78 amino acid sequence has 175 of 309 amino acid residues (56%) identical to, and 222 of 309 amino acid residues (71%) similar to, the 313 amino acid residue ptnr:SPTREMBL-ACC:Q9Z1V0 protein from Mus musculus (Mouse) (OLFACTORY RECEPTOR C6)(. Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV78 is expressed in at least Apical microvilli of the retinal pigment epithelium, arterial (aortic), basal forebrain, brain, Burkitt lymphoma cell lines, corpus callosum, cardiac (atria and ventricle), caudate nucleus, CNS and peripheral tissue, cerebellum, cerebral cortex, colon, cortical neurogenic cells, endothelial (coronary artery and umbilical vein) cells, palate epithelia, eye, neonatal eye, frontal cortex, fetal hematopoietic cells, heart, hippocampus, hypothalamus, leukocytes, liver, fetal liver, lung, lung lymphoma cell lines, fetal lymphoid tissue, adult lymphoid tissue, Those that express MHC II and III nervous, medulla, subthalamic nucleus, ovary, pancreas, pituitary, placenta, pons, prostate, putamen, serum, skeletal muscle, small intestine, smooth muscle (coronary artery in aortic) spinal cord, spleen, stomach, taste receptor cells of the tongue, testis, thalamus, and thymus tissue. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV78 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 78C.

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17474307 ref XP_062466.1 (XM_062466)	similar to olfactory receptor 49 [Homo sapiens]	309	308/309 (99%)	309/309 (99%)	e-141
gi 18480848 gb AAL61438.1 (AY073775)	olfactory receptor MOR115-4 [Mus musculus]	309	238/309 (77%), Positives = 273/309 (88%)	238/309 (77%), Positives = 273/309 (88%)	e-114
gi 18479958 gb AAL60993.1 (AY073330)	olfactory receptor MOR115-1 [Mus musculus]	309	239/309 (77%), Positives = 274/309 (88%)	239/309 (77%), Positives = 274/309 (88%)	e-112
gi 17474309 ref XP_062467.1 (XM_062467)	similar to olfactory receptor 49 [Homo sapiens]	309	247/309 (79%), Positives = 278/309 (89%)	247/309 (79%), Positives = 278/309 (89%)	e-112
gi 18479614 gb AAL60821.1 (AY073158)	olfactory receptor MOR114-1 [Mus musculus]	312	207/306 (67%), Positives = 256/306 (83%)	207/306 (67%), Positives = 256/306 (83%)	e-110

Table 78D lists the domain descriptions from DOMAIN analysis results against NOV78. This indicates that the NOV78 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 78D. Domain Analysis of NOV78

[gnl|Pfam|pfam00001](#), 7tm_1, 7 transmembrane receptor (rhodopsin family).

CD-Length = 254 residues, 100.0% aligned

Score = 90.1 bits (222), Expect = 2e-19

G-Protein Coupled Receptor (GPCRs) have been identified as an extremely large family of protein receptors in a number of species. At the phylogenetic level they can be classified into four major subfamilies. These receptors share a seven transmembrane domain structure with many neurotransmitter and hormone receptors. They are likely to be involved in the recognition and transduction of various signals mediated by G-Proteins, hence their name G-Protein Coupled Receptors. The human GPCR genes are generally intron-less and belong to

10

four gene subfamilies, displaying great sequence variability. These genes are dominantly expressed in olfactory epithelium.

Olfactory receptors (ORs) have been identified as extremely large family of GPCRs in a number of species. As members of the GPCR family, these receptors share a seven transmembrane domain structure with many neurotransmitter and hormone receptors, and are likely to underlie the recognition and G-protein-mediated transduction of odorant signals. Like GPCRs, the ORs they can be expressed in a variety of tissues where they are thought to be involved in recognition and transmission of a variety of signals. The human OR genes are typically intron-less and belong to four different gene subfamilies, displaying great sequence variability. These genes are dominantly expressed in olfactory epithelium.

The disclosed NOV78 nucleic acid of the invention encoding a GPCR-like protein includes the nucleic acid whose sequence is provided in Table 78A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 78A while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 36 percent of the bases may be so changed.

The disclosed NOV78 protein of the invention includes the GPCR-like protein whose sequence is provided in Table 78B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 78B while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 44 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this GPCR-like protein (NOV78) may function as a member of a “GPCR family”. Therefore, the NOV78 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV78 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the GPCR-like protein (NOV78) may be useful in gene therapy, and the GPCR-like protein (NOV78) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from CNS disorders, brain disorders including epilepsy, eating disorders, schizophrenia, ADD; cancer; heart disease; inflammation and autoimmune disorders including Crohn’s disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, blood disorders; psoriasis colon cancer, leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, pancreatic disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer, or other pathologies or conditions. The NOV78 nucleic acid encoding the GPCR-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV78 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV78 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV78 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV79

A disclosed NOV79 nucleic acid of 1601 nucleotides (also referred to as CG57753-01) encoding a GPCR-like protein is shown in Table 79A. The start and stop codons are in bold letters.

5

Table 79A. NOV79 nucleotide sequence (SEQ ID NO:185).

CTGGTTCCTTTGAGTGAGTTATTCCTGGATTCTAGGAGCTCACAGTAGAGTGTTCAGAAT
GGCAAATATCTAAACATTAGCCGGTAATTTTATGCTCCGTATACTGGGTACTAATTTACA
TAAACATATAAGTAAAGTCTACACATATGAGACTGTTTTCTTGATAGATCATGGAAGGAA
AAATCCATTTCAGGGAAAAAAGGGAAATACTATATAAATGTCAAAAATCCAGTCTTTTT
AAGAGACATTCTCTGGAAATATCTCTATTTTGAGGTGTAGTAGATTATCTTACATATATA
TCCACTCACACATACCTTCCAGTTAGAACACTGAAGCCTCATCATTGTAATTAAGCAAT
AAATTTTGTAAAAATGAAAAGGATAATTGTGGGAGGAGATTCTAAACACTCCTTTTCTAA
TGAGCTGCTCTGTGTCGCCAGGGGAAACATGGTTGAGTAAGGCATCACATTTTGTGACATG
GAGCTTCTGACAAATAATCTCAAATTTATCATTGACCCTTTTGTTTACAGGTTCTGACAC
CTTAGTCCAATACCTTCAGAAGAACACATGGAAAATAGGAAAAATTGACTTAATTCATCC
TCTTGGGGCTCACACAGAACCCTGAGGGCCAAAAGTTTATTTGTGCACATTCTTACTCA
TCTACATTGTGACGATAATGGGCAACCTCCTTATCATGGTGACCATCATGGCCAGCCAGT
CCCTGGGTTCCCCCATGTACTTTTTTCTGGCTTCTTATCATTATACATACCGTCTATT
ATACTGCCATTGCTCCCAAATGATTGTTGACCTGCTCTCTGAGAAAAAGACCATTCTTT
TTCAGGGTTGTATGGCTCAACTTTTTATGGATCATTATTTGCTGGTGCTGAGGTCATTCT
TTCTGGTGGTAATGGCCTATGATCAATATGTGGCCATCTGTAAGCCTCTTCATTATTTGA
TCATCATGAATCGTCGAGTCTGTGTTCTCATGCTGTTGGTGGCCTGGATTGGAGGCTTTC
TTCACTCATTGGTTCAATTTCTCTTTATTTATCAGCTCCCTTTCTGTGGACCCAATGTCA
TTGACAACCTTCCTGTGTGATTTGTATCCCTTATTGAACTTGCTTGACCAATACCTATG
TCACTGGGCTTTCTATGATAGCTAATGGTGGAGCGATTTGTACTGTACCTTCTTCCCTC
TCCTGCTTTCTATGGGGTCATATTACCTCTCTTAAGACTCAGAGTTTGAAGGGGAAAT
GCAAAGCTTTCTACACCTGTGCATCCACACTCACTGTGATCACTTTATTCTTTGTCCCT
GCATCTTCTGTTTGTAAAGGCCCACTCCACCTTTCCCATTTGATAAATCCATGCTGG
TTTTAACTTGTATAACTCCCATGCTGAAACCCTAATCTATGCCCTGAGGAATGCAGAAA
TGAAAAGTGCCATGAGGAACTTTGGAGTGAAAAAGTAAGCTTAGCTGGAAGGGCTGT
ATCCCTCATGAGAATATGACTTTCATTCTTTCACAGAAGCAAGGAATAATTTCACTATCC
TATCAGATTACATTTCTGTTATCATTCGCCTTTAGTTATTT

In a search of public sequence databases, the NOV79 nucleic acid sequence, located on chromosome 12 has 605 of 931 bases (64%) identical to a gb:GENBANK-
ID:AF102523|acc:AF102523.1 mRNA from Mus musculus (Mus musculus olfactory receptor
C6 gene, complete cds). Public nucleotide databases include all GenBank databases and the
GeneSeq patent database.

The disclosed NOV79 polypeptide (SEQ ID NO:186) encoded by SEQ ID NO:185 has 277 amino acid residues and is presented in Table 79B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV79 has a signal peptide and is
likely to be localized at the plasma membrane with a certainty of 0.6000. The most likely
cleavage site for a NOV79 peptide is at amino acid position 39.

Table 79B. Encoded NOV79 protein sequence (SEQ ID NO:186).
MGNLLIMVTIMASQSLGSPMYFFLASLSFIHTVYYTAIAPKMIVDLLSEKKTISFQGCMA QLFMDHLFAGAEVILLVVMAYDQYVAICKPLHYLIIMNRRVCVLMLLVAWIGGFLHSLVQ FLFIYQLPFCGPNVIDNFLCDLYPLLKLACTNTYVTGLSMIANGGAICTVTFFPLLLSYG VILPSLKTQSLEGKCKAFYTCASHITVITLFFVPCIFLFVRPNSTFPIDKSMTVVLTTCIT PMLKPLIYALRNAEMKSAMRKLWSEKVSILAGKGLYPS

A search of sequence databases reveals that the NOV79 amino acid sequence has 175
 of 309 amino acid residues (56%) identical to, and 222 of 309 amino acid residues (71%)
 similar to, the 313 amino acid residue ptnr:SPTREMBL-ACC:Q9Z1V0 protein from Mus
 musculus (Mouse) (OLFACTORY RECEPTOR C6)(. Public amino acid databases include the
 5 GenBank databases, SwissProt, PDB and PIR.

NOV79 is expressed in at least Apical microvilli of the retinal pigment epithelium,
 arterial (aortic), basal forebrain, brain, Burkitt lymphoma cell lines, corpus callosum, cardiac
 (atria and ventricle), caudate nucleus, CNS and peripheral tissue, cerebellum, cerebral cortex,
 10 colon, cortical neurogenic cells, endothelial (coronary artery and umbilical vein) cells, palate
 epithelia, eye, neonatal eye, frontal cortex, fetal hematopoietic cells, heart, hippocampus,
 hypothalamus, leukocytes, liver, fetal liver, lung, lung lymphoma cell lines, fetal lymphoid
 tissue, adult lymphoid tissue, Those that express MHC II and III nervous, medulla,
 subthalamic nucleus, ovary, pancreas, pituitary, placenta, pons, prostate, putamen, serum,
 15 skeletal muscle, small intestine, smooth muscle (coronary artery in aortic) spinal cord, spleen,
 stomach, taste receptor cells of the tongue, testis, thalamus, and thymus tissue. This
 information was derived by determining the tissue sources of the sequences that were included
 in the invention including but not limited to SeqCalling sources, Public EST sources,
 Literature sources, and/or RACE sources.

20 The disclosed NOV79 polypeptide has homology to the amino acid sequences shown
 in the BLASTP data listed in Table 79C.

Table 79C. BLAST results for NOV79					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17459952 ref XP_062090.1 (XM_062090)	similar to odorant receptor 16 [Homo sapiens]	277	275/277 (99%)	276/277 (99%)	e-133
gi 17460091 ref XP_062159.1 (XM_062159)	similar to odorant receptor 16 [Homo sapiens]	277	275/277 (99%)	275/277 (99%)	e-132

<u>gi 17460099 ref XP_062161.1 </u> (XM_062161)	similar to odorant receptor 16 [Homo sapiens]	722	242/266 (90%)	251/266 (93%)	e-119
<u>gi 18479528 gb AAL60778.1 </u> (AY073115)	olfactory receptor MOR231-2 [Mus musculus]	314	226/277 (81%)	247/277 (88%)	e-109
<u>gi 18479534 gb AAL60781.1 </u> (AY073118)	olfactory receptor MOR231-3 [Mus musculus]	305	186/265 (70%)	223/265 (83%)	7e-94

Table 79D lists the domain descriptions from DOMAIN analysis results against NOV79. This indicates that the NOV79 sequence has properties similar to those of other proteins known to contain this domain.

5

<p align="center">Table 79D. Domain Analysis of NOV79</p> <p align="center"><u>gnl Pfam pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family).</u></p> <p align="center">CD-Length = 254 residues, 100.0% aligned</p> <p align="center">Score = 91.3 bits (225), Expect = 7e-20</p>	
--	--

G-Protein Coupled Receptor (GPCRs) have been identified as an extremely large family of protein receptors in a number of species. At the phylogenetic level they can be classified into four major subfamilies. These receptors share a seven transmembrane domain structure with many neurotransmitter and hormone receptors. They are likely to be involved in the recognition and transduction of various signals mediated by G-Proteins, hence their name G-Protein Coupled Receptors. The human GPCR genes are generally intron-less and belong to four gene subfamilies, displaying great sequence variability. These genes are dominantly expressed in olfactory epithelium.

10

15

20

Olfactory receptors (ORs) have been identified as extremely large family of GPCRs in a number of species. As members of the GPCR family, these receptors share a seven transmembrane domain structure with many neurotransmitter and hormone receptors, and are likely to underlie the recognition and G-protein-mediated transduction of odorant signals. Like GPCRs, the ORs they can be expressed in a variety of tissues where they are thought to be involved in recognition and transmission of a variety of signals. The human OR genes are typically intron-less and belong to four different gene subfamilies, displaying great sequence variability. These genes are dominantly expressed in olfactory epithelium. The disclosed NOV79 nucleic acid of the invention encoding a GPCR-like protein includes the nucleic acid whose sequence is provided in Table 79A or a fragment thereof. The invention also includes a

mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 79A while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including
 5 nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the
 10 chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 36 percent of the bases may be so changed.

The disclosed NOV79 protein of the invention includes the GPCR-like protein whose
 15 sequence is provided in Table 79B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 79B while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 44 percent of the residues may be so changed.

20 The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this GPCR-like protein (NOV79) may function as a member of a "GPCR family". Therefore, the NOV79 nucleic acids and proteins identified here may be useful in potential therapeutic applications
 25 implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues
 30 and cell types composing (but not limited to) those defined here.

The NOV79 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the GPCR-like protein (NOV79) may be useful in gene therapy, and the GPCR-like protein (NOV79) may be useful

CCCAAGCCATTGATTCCCAGAGCTCCAACAAGCTCATCTCTGCCGTGTACACTGTTGTACGCCAATAAT
TAACCCTTTGATTTACTGCCTGAGGAACAAGGAATTAAGGACGCCTTGAAAAAGGCCTTGGGCTTGGGT
CAACTTCACACTAAGACAATAAT

The disclosed NOV80 polypeptide (SEQ ID NO:188) encoded by SEQ ID NO:187 has 324 amino acid residues and is presented in Table 80B using the one-letter amino acid code.

Table 80B. Encoded NOV80 protein sequence (SEQ ID NO:188).

MFCRPAAPKHRGMSGENVTKVSTFILVGLPTAPGLQYLLFLLFLLTYLFLVLENLAAILIVWSSTSLHRP
MYFYLSSMSFLEIWYVSDITPKMLEGFLQKRI SFVGCMTQLYFFSSLVCTECVLLASMAYDRYVAICH
PLRYHVLVTPGLCLQLVGFSFVSGFTISMIKVCFISSVTFCSNVLNHHFFCDISPILKLACTDFSTAEIV
DFILAFIILVFPLLATILSYWHITLAVLRIPSATGCWRAFSTCASHLTVVTVFYTALLFMYVRPQAIDSQ
SSNKLISAVYTVVTPINPLIYCLRNKEFKDALKKALGLGQTS

A search of sequence databases reveals that the NOV80 amino acid sequence has 215/305 (70%) identity and 253/305 (82%) similarity with TREMBLNEW-ACC:AAG45189 M51 OLFACTORY RECEPTOR - Mus musculus. Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

The disclosed NOV80 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 80C.

Table 80C. BLAST results for NOV80

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17435888 ref XP_065377.1 (XM_065377)	similar to olfactory receptor 41 [Homo sapiens]	312	312/312 (100%)	312/312 (100%)	e-142
gi 17435880 ref XP_065375.1 (XM_065375)	similar to olfactory receptor 41 [Homo sapiens]	331	293/307 (95%)	300/307 (97%)	e-132
gi 18479396 gb AAL60712.1 (AY073049)	olfactory receptor MOR103-2 [Mus musculus]	312	271/309 (87%)	289/309 (92%)	e-123
gi 18479398 gb AAL60713.1 (AY073050)	olfactory receptor MOR103-3 [Mus musculus]	312	268/310 (86%)	288/310 (92%)	e-123
gi 12007416 gb AAG45189.1 (AF321234)	m51 olfactory receptor [Mus musculus]	314	215/305 (70%)	253/305 (82%)	e-102

Table 80D lists the domain descriptions from DOMAIN analysis results against NOV80. This indicates that the NOV80 sequence has properties similar to those of other proteins known to contain this domain.

Table 80D. Domain Analysis of NOV80

gnl|Pfam|pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family).
 CD-Length = 254 residues, 100.0% aligned
 Score = 98.6 bits (244), Expect = 5e-22

The disclosed NOV80 nucleic acid of the invention encoding a GPCR-like protein includes the nucleic acid whose sequence is provided in Table 80A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 80A while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 5 percent of the bases may be so changed.

The disclosed NOV80 protein of the invention includes the GPCR-like protein whose sequence is provided in Table 80B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 80B while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 30 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this GPCR-like protein (NOV80) may function as a member of a "GPCR family". Therefore, the NOV80 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein

therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

5 The NOV80 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the GPCR-like protein (NOV80) may be useful in gene therapy, and the GPCR-like protein (NOV80) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the
10 compositions of the present invention will have efficacy for treatment of patients suffering from CNS disorders, brain disorders including epilepsy, eating disorders, schizophrenia, ADD; cancer; heart disease; inflammation and autoimmune disorders including Crohn's disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, blood disorders; psoriasis colon cancer, leukemia AIDS; thalamus disorders; metabolic disorders including
15 diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, pancreatic disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer, or other pathologies or conditions. The NOV80 nucleic acid encoding the GPCR-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be
20 assessed.

NOV80 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV80 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-
25 NOVX Antibodies" section below. The disclosed NOV80 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

30 **NOV81**

NOV81 includes two GPCR-like proteins disclosed below. The disclosed sequences have been named NOV81a and NOV81b.

NOV81a

A disclosed NOV81a nucleic acid of 1039 nucleotides (also referred to as CG57847-01) encoding a GPCR-like protein is shown in Table 81A. The start and stop codons are in bold letters.

5

Table 81A. NOV81a nucleotide sequence (SEQ ID NO:189).
<p> GAATGATGCCCTTTTGCCACAATATAATTAATATTTCTGTGTGAAAAACAACTGGTCAAATGATGTCCG TGCTTCCTGTACAGTTTAATGGTGCTCATAATTCGACCACACTCGTTGGCAATCTGATAGTTATTGTT TCTATATCACACTTCAAACAACCTCATACCCCAACAAATTGGCTCATTCCATGGCCACTGTGGACT TTCTTCTGGGGTGTCTGGTCATGCCCTTACAGTATGGTGAGATCTGCTGAGCACTGTTGGTATTTTGGAGA AGTCTTCTGTAAAATTACACAAGCACCGACATTATGCTGAGCTCAGCCTCCATTTCCATTTGTCTTTC ATCTCCATTGACCGCTACTATGCTGTGTGTGATCCACTGAGATATAAAGCCAAGATGAATATCTTGGTTA TTTGTGTGATGATCTTCATTAGTTGGAGTGTCCTGCTGTTTTTGCATTTGGAATGATCTTCTGGAGCT AAAGTTCAAAGGCGCTGAAGAGATATATTACAAACATGTTCACTGCAGAGGAGGTTGCTCTGTCTCTTT AGCAAAATATCTGGGTTACTGACCTTTATGACTTCTTTTATATACCTGGATCTATTATGTTATGTGTCT ATTACAGAATATATCTTATCGCTAAAGAACAGGCAAGATTAATTAGTGATGCCAATCAGAAGCTCCAAAT TGGATTGGAATGAAAAATGGAATTTACAAAGCAAAGAAAGGAAAGCTGTGAAGACATTGGGGATTGTG ATGGGAGTTTTTCTAATATGCTGGTGCCCTTCTTTATCTGTACAGTCATGGACCCTTTTCTTCACTACA TTATTCCACCTACTTTGAATGATGTATTGATTGGTTTGGCTACTTGAAGCTCTACATTTAATCCAATGGT TTATGCATTTTTCTATCCTTGGTTTAGAAAAGCACTGAAGATGATGCTGTTTGGTAAAATTTTCCAAAA GATTCATCCAGGTGTAATATTATTTTGAATTGAGTTCATAGAATTATTATATTTTACT </p>

The disclosed NOV81a polypeptide (SEQ ID NO:190) encoded by SEQ ID NO:189 has 339 amino acid residues and is presented in Table 81B using the one-letter amino acid code.

10

Table 81B. Encoded NOV81a protein sequence (SEQ ID NO:190).
<p> MMPFCHNIINISCVKNNSNDVRASLYSLMVLIIILTTLVGNLIVIVSISHFKQLHPTPNWLIHSMATVDF LLGCLVMPYSMVRSAEHCWYFGEVFCKIHTSTDIMLSSASIFHLSFISIDRYAVCDPLRYKAKMNILVI CVMIFISWSVPAVFAFGMIFLELNFKGAEIYYKHVHCRGGCSVFFSKISGVLTFMTSFYIPGSIMLCVY YRIYLIAKEQARLISDANQKLQIGLEMKNGISQSKERKAVKTLGIVMGVFLICWCPFFICTVMDPFLHYI IPPTLNDVLIWFGYLNSTFNPVMVYAFFYPWFRKALKMMLFGKIFQKDSSRCKLFLELSS </p>

A search of sequence databases reveals that the NOV81a amino acid sequence has 152/299 (50%) identity and 206/299 (68%) similarity to SPTREMBL-ACC:Q9P1P4 G PROTEIN-COUPLED RECEPTOR 57 - Homo sapiens. Public amino acid databases include 15 the GenBank databases, SwissProt, PDB and PIR.

NOV81b

A disclosed NOV81b nucleic acid of 1039 nucleotides (also referred to as CG57847-02) encoding a GPCR-like protein is shown in Table 81C. The start and stop codons are in 20 bold letters.

Table 81C. NOV81b nucleotide sequence (SEQ ID NO:191).

GAATGATGCCCTTTTGCCACAATATAATTAATATTTTCCTGTGTGAAAAACAACCTGGTCAA
ATGATGTCCGTGCTTCCCTGTACAGTTTAATGGTGCTCATAATTCTGACCACACTCGTTG
GCAATCTGATAGTTATTGTTTCTATATCACACTTCAAACAACCTCATACCCAACAAATT
GGCTCATTCCATGGCCACTGTGGACTTTCTTCTGGGGTGTCTGGTCATGCCTTACA
GTATGGTGAGATCTGCTGAGCACTGTTGGTATTTTGGAGAAGTCTTCTGTAAAATTACACA
CAAGCACCGACATTATGCTGAGCTCAGCCTCCATTTTCCATTTGTCTTTCATCTCCATTG
ACCGTACTATGCTGTGTGTGATCCACTGAGATATAAAGCCAAGATGAATATCTTGGTTA
TTTGTGTGATGATCTTCATTAGTTGGAGTGTCCTGCTGTTTTTGCATTTGGAATGATCT
TTCTGGAGCTAAACTTCAAAGGCGCTGAAGAGATATATTACAAACATGTTCACTGCAGAG
GAGGTTGCTCTGTCTTCTTTAGCAAAATATCTGGGGTACTGACCTTTATGACTTCTTTTT
ATATACCTGGATCTATTATGTTATGTGTCTATTACAGAATATATCTTATCGCTAAAGAAC
AGGCAAGATTAATTAGTGATGCCAATCAGAAGCTCCAAATTGGATTGGAAATGAAAAATG
GAATTTACAAAGCAAAGAAAGGAAAGCTGTGAAGACATTGGGGATTGTGATGGGAGTTT
TCCTAATATGCTGGTGCCCTTTCTTTATCTGTACAGTCATGGACCCTTTCTTCACTACA
TTATTCACCTACTTTGAATGATGTATTGATTGGTTTGGCTACTTGAACCTCTACATTTA
ATCCAATGGTTTATGCATTTTTCTATCCTTGGTTTAGAAAAGCACTGAAGATGATGCTGT
TTGGTAAAATTTTCCAAAAGATTCATCCAGGTGTAAATTATTTTGGAAATTGAGTTCAT
AGAATTATTATTTTACT

In a search of public sequence databases, the NOV81b nucleic acid sequence, located on chromosome 6 has 616 of 979 bases (62%) identical to a gb:GENBANK-

- 5 ID:HSU88828|acc:U88828.1 mRNA from Homo sapiens (Homo sapiens serotonin-4-receptor-like pseudogene). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV81b polypeptide (SEQ ID NO:192) encoded by SEQ ID NO:191 has 339 amino acid residues and is presented in Table B using the one-letter amino acid code.

- 10 Signal P, Psort and/or Hydropathy results predict that NOV81b has a signal peptide and is likely to be localized in the plasma membrane with a certainty of 0.6000. The most likely cleavage site for a NOV81b peptide is between amino acids 47 and 48.

Table 81D. Encoded NOV81b protein sequence (SEQ ID NO:192).

MMPFCHNIINISCVKNNSNDVRSALYSLMVLIIITTLVGNLIVIVSISHFKQLHTPTNW
LIHSMATVDFLLGCLVMPYSMVRSAEHCWYFGEVFCIKHTSTDIMLSSASIFHLSFISID
RYYAVCDPLRYKAKMNILVICVMIFISWSVPAVFAFGMIFLELNFKGAEIYYKHVHCRG
GCSVFFSKISGVLTFMTSFYIPGSIMLCVYYRIYLIAKEQARLISDANQKLQIGLEMKNG
ISQSKERKAVKTLGIVMGVFLICWCPFFICTVMDPFLHYIIPPTLNDVLIWFGYLNSTFN
PMVYAFFYPWFRKALKMMLFGKIFQKDSSRCKLFLELSS

- 15 A search of sequence databases reveals that the NOV81b amino acid sequence has 152 of 299 amino acid residues (50%) identical to, and 206 of 299 amino acid residues (68%) similar to, the 343 amino acid residue ptrn:SPTREMBL-ACC:Q9P1P4 protein from Homo sapiens (Human) (G PROTEIN-COUPLED RECEPTOR 57). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV81b is expressed in at least Apical microvilli of the retinal pigment epithelium, arterial (aortic), basal forebrain, brain, Burkitt lymphoma cell lines, corpus callosum, cardiac (atria and ventricle), caudate nucleus, CNS and peripheral tissue, cerebellum, cerebral cortex, colon, cortical neurogenic cells, endothelial (coronary artery and umbilical vein) cells, palate epithelia, eye, neonatal eye, frontal cortex, fetal hematopoietic cells, heart, hippocampus, hypothalamus, leukocytes, liver, fetal liver, lung, lung lymphoma cell lines, fetal lymphoid tissue, adult lymphoid tissue, Those that express MHC II and III nervous, medulla, subthalamic nucleus, ovary, pancreas, pituitary, placenta, pons, prostate, putamen, serum, skeletal muscle, small intestine, smooth muscle (coronary artery in aortic) spinal cord, spleen, stomach, taste receptor cells of the tongue, testis, thalamus, and thymus tissue. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV81b polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 81E.

Table 81E. BLAST results for NOV81b					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 10441577 gb AAG17112.1 AF200627.1 (AF200627)	putative catecholamine receptor [Homo sapiens]	339	339/339 (100%)	339/339 (100%)	e-177
gi 17453976 ref XP_069048.1 (XM_069048)	similar to trace amine receptor 1 [Homo sapiens]	338	338/338 (100%)	338/338 (100%)	e-176
gi 14600076 gb AAK71237.1 AF380186.1 (AF380186)	trace amine receptor 1 [Rattus norvegicus]	332	261/334 (78%), Positives = 288/334 (86%),	261/334 (78%), Positives = 288/334 (86%),	e-136
gi 18182341 gb AAL65137.1 AF421352.1 (AF421352)	trace amine receptor 1 [Rattus norvegicus]	332	261/334 (78%), Positives = 287/334 (85%),	261/334 (78%), Positives = 287/334 (85%),	e-136
gi 16716513 ref NP_444435.1 (NM_053205)	trace amine receptor 1 [Mus musculus]	332	252/334 (75%), Positives = 283/334 (84%),	252/334 (75%), Positives = 283/334 (84%),	e-133

Table 81F lists the domain descriptions from DOMAIN analysis results against NOV81b. This indicates that the NOV81b sequence has properties similar to those of other proteins known to contain this domain.

therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

5 The NOV81b nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the GPCR-like protein (NOV81b) may be useful in gene therapy, and the GPCR-like protein (NOV81b) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the
10 compositions of the present invention will have efficacy for treatment of patients suffering from CNS disorders, brain disorders including epilepsy, eating disorders, schizophrenia, ADD; cancer; heart disease; inflammation and autoimmune disorders including Crohn's disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, blood disorders; psoriasis colon cancer, leukemia AIDS; thalamus disorders; metabolic disorders including
15 diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, pancreatic disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer, or other pathologies or conditions. The NOV81b nucleic acid encoding the GPCR-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be
20 assessed.

 NOV81b nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV81b substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-
25 NOVX Antibodies" section below. The disclosed NOV81b proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

30 **NOV82**

 A disclosed NOV82 nucleic acid of 1033 nucleotides (also referred to as CG57845-01) encoding a GPCR-like protein is shown in Table 82A. The start and stop codons are in bold letters.

Table 82A. NOV82 nucleotide sequence (SEQ ID NO:193).

AACCATGACCAGCAATTTTCCCAACCTGTTGTGCAGCTTTGCTATGAGGATGTGAATGGATCTTGTATT
GAAACTCCCTATTCTCCTGGGTCCTGGGTAATTCGTACACGGCGTTTAGCTTTGGGCTTTGCTGGCTG
TATTTGGAAATCTCTTAGTAATGACTTCTGTTCTTCAATTTAAGCAGCTGCACTCTCCAACCAATTTTCT
CATTGCCTCTCTGGCCTGTGCTGACTTCTTGGTAGGTGTGACTGTGATGCTTTTCAGCATGGTCAGGACG
GTGGAGAGCTGCTGGTATTTTGGAGCCAAATTTTGTACTCTTCACAGTTGCTGTGATGTGGCATTTTGTT
ACTCTTCTGTCCTCCACTTGTGCTTCATCTGCATCGACAGGTACATTGTGGTTACTGATCCCTGGTCTA
TGCTACCAAGTTCACCGTGTCTGTGTCGGGAATTTGCATCAGCGTGTCTGGATTCTGCCTCTCACGTAC
AGCGGTGCTGTGTTCTACACAGGTGTCAATGATGATGGGCTGGAGGAATTAGTAAGTGCTCTCAACTGCG
TAGGTGGCTGTCAAATATTGTAAGTCAAGGCTGGGTGTTGATAGATTTTCTGTTATTCTTCATACCTAC
CCTTGTATGATAATTCTTTACAGTAAGATTTTCTTATAGCTAAACAACAAGCTATAAAAATTGAACT
ACTAGTAGCAAAGTAGAATCATCCTCAGAGAGTTATAAAATCAGAGTGGCCAAGAGAGAGAGGAAAGCAG
CTAAACCTGGGGGTACCGTACTAGCATTTGTTATTTTATGTTACCGTATACAGTTGATATATTAAT
TGATGCCTTTATGGGCTTCTGACCCCTGCCTATATCTATGAAATTTGCTGTGGAGTGCTTATTATAAC
TCAGCCATGAATCCTTTGATTTATGCTCTATTTTATCCTTGGTTAGGAAAGCCATAAACTATTTTAA
GTGGAGATGTTTTAAAGGCTAGTTCATCAACCATTAGTTTATTTTAGAATAA

The disclosed NOV82 polypeptide (SEQ ID NO:194) encoded by SEQ ID NO:193 has 342 amino acid residues and is presented in Table 82B using the one-letter amino acid code.

5

Table 82B. Encoded NOV82 protein sequence (SEQ ID NO:194).

MTSNFSQPVVQLCYEDVNGSCIETPYPSPGSRVILYAFSFGSLLAVFGNLLVMTSVLHFKQLHSPTNFLI
ASLACADFLVGVTVMFLFSMVRTVESWCYFGAKFCTLHSCCDVAFYSSVLHLCFICIDRYIVVTDPLVYA
TKFTVSVSGICISVSWILPLTYSGAVFYTGTVNDGLEELVSALNCVGGCQIIIVSQGWVLIDFLFFIPTL
VMIIILYSKIFLIAKQQAIIKIETSSKVESSESSEYKIRVAKRERKAAKTLGVTVLAFVISWLPYTVTDILID
AFMGFLTPAYIYEICWSAYYNSAMNPLIYALFPWFRKAIIKLILSGDVLKASSSTISLFLE

A search of sequence databases reveals that the NOV82 amino acid sequence has 145/330 (43%) identity and 214/330 (64%) similarity with SPTREMBL-ACC:O14804 PUTATIVE NEUROTRANSMITTER RECEPTOR - Homo sapiens. Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

10

The disclosed NOV82 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 82C.

Table 82C. BLAST results for NOV82

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 16751917 ref NP_444508.1 (NM_053278)	G protein-coupled receptor 102; trace amine receptor 5 [Homo sapiens]	342	342/342 (100%)	342/342 (100%)	e-166
gi 14600102 gb AAK71250.1 AF380199_1 (AF380199)	similar to trace amine receptor 4 (H. sapiens) [Homo sapiens]	374	274/344 (79%)	308/344 (88%)	e-143

<u>gi 17453968 ref XP_069046.1 </u> (XM_069046)	Similar to trace amine receptor 4 [Rattus norvegicus]	345	273/343 (79%)	303/343 (87%)	e-142
<u>gi 14600086 gb AAK71242.1 AF380191.1</u> (AF380191)	trace amine receptor 4 [Rattus norvegicus]	345	266/343 (77%)	302/343 (87%)	e-142
<u>gi 14600100 gb AAK71249.1 AF380198.1</u> (AF380198)	trace amine receptor 10 [Rattus norvegicus]	344	274/344 (79%)	305/344 (88%)	e-141

Table 82D lists the domain descriptions from DOMAIN analysis results against NOV82. This indicates that the NOV82 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 82D. Domain Analysis of NOV82	
<u>gnl Pfam pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family).</u>	CD-Length = 254 residues, 100.0% aligned
	Score = 132 bits (331), Expect = 4e-32

The disclosed NOV82 nucleic acid of the invention encoding a GPCR-like protein includes the nucleic acid whose sequence is provided in Table 82A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 82A while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 5 percent of the bases may be so changed.

The disclosed NOV82 protein of the invention includes the GPCR-like protein whose sequence is provided in Table 82B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 82B

while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 57 percent of the residues may be so changed.

5 The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

10 The above defined information for this invention suggests that this GPCR-like protein (NOV82) may function as a member of a “GPCR family”. Therefore, the NOV82 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

15 The NOV82 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the GPCR-like protein (NOV82) may be useful in gene therapy, and the GPCR-like protein (NOV82) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from CNS disorders, brain disorders including epilepsy, eating disorders, schizophrenia, ADD; cancer; heart disease; inflammation and autoimmune disorders including Crohn’s disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, blood disorders; psoriasis colon cancer, leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, pancreatic disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer, or other pathologies or conditions. The NOV82 nucleic acid encoding the GPCR-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

20

25

30

NOV82 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV82 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-

NOV83

A disclosed NOV83 nucleic acid of 1045 nucleotides (also referred to as CG57843-01) encoding a GPCR-like protein is shown in Table 83A. The start and stop codons are in bold letters.

Table 83A. NOV83 nucleotide sequence (SEQ ID NO:195).

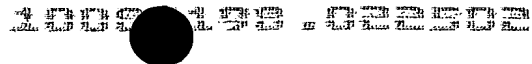
CGTTATGAGCAGCAATTTCATCCCTGCTGGTGGCTGTGCAGCTGTGCTACGCGAACGTGAATGGGTCCTGT
GTGAAAATCCCTTCTCGCCGGGATCCCGGGTGATTCTGTACATAGTGTGGCTTTGGGGCTGTGCTGG
CTGTGTTTGAAAACCTCCTGGTGATGATTTCAATCCTCCATTTCAGCAGCTGCATCTCCGACCAATT
TCTCGTTGCCCTCTCGGCCCTCGCTGATTTCTGGTGGGTGTGACTGTGATGCCCTTCAGCATGGTTCAGG
ACGGTGGAGAGCTGCTGGTATTTTGGGAGGAGTTTTGTACTTTCCACACCTGTGTGATGTGGCATTTT
GTTACTCTTCTCTCTTTCACTTGTGCTTCATCTCCATCGACAGGTACATTGCGGTTACTGACCCCTGGT
CTATCCTACCAAGTTCACCGTATCTGTGTCAGGAATTTGCATCAGCGTGTCTGGATCCTGCCCCTCATG
TACAGCGGTGTGTGTTTCTACACAGGTGTCTATGACGATGGGCTGGAGGAATTATCTGATGCCCTAAACT
GTATAGGAGGTTGTGCAGACCTGTGTAATCAAAACTGGGTGTGACAGATTTCTATCTTCTTTATACC
TACCTTTATATGATAATCTGTATGGTAACATATTTCTTGTGGCTAGACGACAGGCGAAAAAGATAGAA
AATACTGGTAGCAAGACAGAATCATCCTCAGAGAGTTACAAAGCCAGAGTGGCCAGGAGAGAGAGAAAAAG
CAGCTAAAACCTGGGGGTACAGTGGTAGCATTTATGATTTTCATGGTTACCATATAGCATTGATTCATT
AATTGATGCCCTTTATGGGCTTTATAACCCCTGCCTGTATTTATGAGATTTGCTGTTGGTGTGCTTATTAT
AACTCAGCCATGAATCCTTTGATTTATGCTTTATTTACCATGGGTTAGGAAAAGCAATAAAGTATTATG
TAAGTGGTCAGGTTTTAAAGAACAGTTTCAGCAACCATGAATTTGTTTTCTGAACATATATAAGCA

The disclosed NOV83 polypeptide (SEQ ID NO:196) encoded by SEQ ID NO:195 has 345 amino acid residues and is presented in Table 83B using the one-letter amino acid code.

Table 83B. Encoded NOV83 protein sequence (SEQ ID NO:196).

MSSNSSLLVAVQLCYANVNGSCVKIPFSPGSRVILYIVFGFGAVLAVFGNLLVMISILHFKQLHSPTNFL
 VASLACADFLVGVTVMPFPMVRTVESCWYFGRSFCTFHTCCDVAFCYSSLFHLCFISIDRYIAVTDPLVY
 PTKFTVSVSGICISVSWILPLMYSYGAVFYTGVDGDEELSDALNCGGCCQTVVNQNWVLTDPLSFSPFIPT
 FIMILYGNILFVARQKAKIENTGSKTSSSESYKARVARRERKAAKTLGVTVVFAMISWLPYSIDSLI
 DAFMGFITPACTYEICCCWACYNSAMNPLIYALFYPPWFRKAIKVIVTGQVLKNSSATMNLFSFHSI

A search of sequence databases reveals that the NOV83 amino acid sequence has 146/330 (44%) identity and 216/330 (65%) similarity with SPTREMBL-ACC:O14804 PUTATIVE NEUROTRANSMITTER RECEPTOR - Homo sapiens. Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.



The disclosed NOV83 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 83C.

Table 83C. BLAST results for NOV83					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
<u>gi 17453968 ref XP_069046.1 </u> (XM_069046)	similar to trace amine receptor 4 (H. sapiens)	345	345/345 (100%)	345/345 (100%)	e-180
<u>gi 14600086 gb AAK71242.1 AF380191.1</u> (AF380191)	trace amine receptor 4 [Rattus norvegicus]	345	302/345 (87%)	321/345 (92%)	e-162
<u>gi 14600102 gb AAK71250.1 AF380199.1</u> (AF380199)	trace amine receptor 11 [Rattus norvegicus]	373	265/345 (76%)	306/345 (87%)	e-141
<u>gi 16751917 ref NP_444508.1 </u> (NM_053278)	G protein-coupled receptor 102; trace amine receptor 5 [Homo sapiens]	342	273/343 (79%)	303/343 (87%)	e-137
<u>gi 14600094 gb AAK71246.1 AF380195.1</u> (AF380195)	trace amine receptor 7 [Rattus norvegicus]	344	258/345 (74%)	302/345 (86%)	e-137

5 Table 83D lists the domain descriptions from DOMAIN analysis results against NOV83. This indicates that the NOV83 sequence has properties similar to those of other proteins known to contain this domain.

Table 83D. Domain Analysis of NOV83	
<u>gnl Pfam pfam00001</u> , 7tm_1, 7 transmembrane receptor (rhodopsin family).	CD-Length = 254 residues, 100.0% aligned
Score = 130 bits (326), Expect = 2e-31	

10 The disclosed NOV83 nucleic acid of the invention encoding a GPCR-like protein includes the nucleic acid whose sequence is provided in Table 83A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 83A while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a fragment of such a nucleic acid. The
15 invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or



complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 5 percent of the bases may be so changed.

The disclosed NOV83 protein of the invention includes the GPCR-like protein whose sequence is provided in Table 83B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 83B while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 56 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this GPCR-like protein (NOV83) may function as a member of a “GPCR family”. Therefore, the NOV83 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV83 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the GPCR-like protein (NOV83) may be useful in gene therapy, and the GPCR-like protein (NOV83) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from CNS disorders, brain disorders including epilepsy, eating disorders, schizophrenia, ADD; cancer; heart disease; inflammation and autoimmune disorders including Crohn’s disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, blood disorders; psoriasis colon cancer, leukemia AIDS; thalamus disorders; metabolic disorders including

The disclosed NOV84a polypeptide (SEQ ID NO:198) encoded by SEQ ID NO:197 has 312 amino acid residues and is presented in Table 84B using the one-letter amino acid code.

Table 84B. Encoded NOV84a protein sequence (SEQ ID NO:198).

MALGNHSTITEFLLLGLSADPNIRALLFVLFLGIYLLTIMENLMLLLMIRADSCILHKPMYFFLSHLSFVD
 LCFSSVIVPKMLENLLSQRKTIISVEGCLAQVFFVFVFTAGTEACLLSGMAYDRHAAICRPLLYGQIMGKQL
 YMHLVWGSWGLGFLDALINVLLAVNMVFCIAKI IHHSYEMPSLLPLSCSDISRSLIALLCSTLLHGLGN
 FLLVFLSYTRIISTILSISSTSGRSKAFSTCSAHLTAVTLYYGSGLLRHLPNSGSPIELIFSVQYTVVT
 PMLNSLIYSLKNKEVKVALKRTLEKYLQYTRR

A search of sequence databases reveals that the NOV84a amino acid sequence has 138/309 (44%) identity and 203/309 (65%) similarity with TREMBLNEW-ACC:AAG39860 ODORANT RECEPTOR K15 - Mus musculus. Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV84b

A disclosed NOV84b nucleic acid of 1039 nucleotides (also referred to as CG57841-02) encoding a GPCR-like protein is shown in Table 84C. The start and stop codons are in bold letters.

Table 84C. NOV84b nucleotide sequence (SEQ ID NO:199).

GAATGATGCCCTTTTGCCACAATATAATTAATATTTCTGTGTGAAAAACAACTGGTCAA
 ATGATGTCCGTGCTTCCCTGTACAGTTTAATGGTGCTCATAATTCGACCACACTCGTTG
 GCAATCTGATAGTTATTGTTTCTATATCACACTTCAAACAACCTTCATACCCCAACAAATT
 GGCTCATTCCATCCATGGCCACTGTGGACTTTCTTCTGGGGTGTCTGGTCATGCCTTACA
 GTATGGTGAGATCTGCTGAGCACTGTTGGTATTTTGGAGAAGTCTTCTGTAAAATTCACA
 CAAGCACCGACATTATGCTGAGCTCAGCCTCCATTTTCCATTTGTCTTTCATCTCCATTG
 ACCGCTACTATGCTGTGTGTGATCCACTGAGATATAAAGCCAAGATGAATATCTTGGTTA
 TTTGTGTGATGATCTTCATTAGTTGGAGTGTCCCTGCTGTTTTTGCATTTGGAATGATCT
 TTCTGGAGCTAAACTTCAAAGGCGCTGAAGAGATATATTACAAACATGTTCACTGCAGAG
 GAGGTTGCTCTGTCTTCTTTAGCAAAATATCTGGGGTACTGACCTTTATGACTTCTTTTT
 ATATACCTGGATCTATTATGTTATGTGTCTATTACAGAATATATCTTATCGCTAAAGAAC
 AGGCAAGATTAATTAGTGATGCCAATCAGAAGCTCCAAATTGGATTGGAATGAAAAATG
 GAATTTACAAAGCAAAGAAAGGAAAGCTGTGAAGACATTGGGGATTGTGATGGGAGTTT
 TCCTAATATGCTGGTGCCCTTTCTTTATCTGTACAGTCATGGACCCTTTTCTTCACTACA
 TTATTCACCTACTTTGAATGATGTATTGATTTGGTTTGGCTACTTGAACCTTACATTTA
 ATCCAATGGTTTATGCATTTTCTATCCTTGTTTGAAGAAAGCACTGAAGATGATGCTGT
 TTGGTAAAATTTTCCAAAAAGATTCATCCAGGTGTAATTTATTTTGGGAATTGAGTTCAT
AGAATTATTATATTTTACT

In a search of public sequence databases, the NOV84b nucleic acid sequence, located on chromosome 6 has 616 of 979 bases (62%) identical to a gb:GENBANK-ID:HSU88828|acc:U88828.1 mRNA from Homo sapiens (Homo sapiens serotonin-4-receptor-

like pseudogene). Public nucleotide databases include all GenBank databases and the GeneSeq patent database.

The disclosed NOV84b polypeptide (SEQ ID NO:200) encoded by SEQ ID NO:199 has 339 amino acid residues and is presented in Table 84D using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV84b has a signal peptide and is likely to be localized in the plasma membrane with a certainty of 0.6000. The most likely cleavage site for a NOV84b peptide is between amino acids 47 and 48.

Table 84D. Encoded NOV84b protein sequence (SEQ ID NO:200).

MMPFCHNIINISCVKNNWSNDVRASLYSLMVLIIILTTLVGNLIVIVSISHFKQLHTPTNW
LIHSMATVDFLGCLVMPYSMVRSAEHCWYFGEVFCIKHTSTDIMLSSASIFHLSFISID
RYYAVCDPLRYKAKMNLVICVMIFISWSVPAVFAFGMIFLELNFKGAEEIYYKHVHCRG
GCSVFFSKISGVLTFMTSFYIPGSIMLCVYYRIYLIQAEQARLISDANQKLQIGLEMKNG
ISQSKERKAVKTLGIVMGVFLICWCPFFICTVMDPFLHYIIPPTLNDVLIWFGYLNSTFN
PMVYAFFYPWFRKALKMMLFGKIFQKDSSRCKLFLELSS

A search of sequence databases reveals that the NOV84b amino acid sequence has 152 of 299 amino acid residues (50%) identical to, and 206 of 299 amino acid residues (68%) similar to, the 343 amino acid residue ptnr:SPTREMBL-ACC:Q9P1P4 protein from Homo sapiens (Human) (G PROTEIN-COUPLED RECEPTOR 57). Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

NOV84b is expressed in at least Apical microvilli of the retinal pigment epithelium, arterial (aortic), basal forebrain, brain, Burkitt lymphoma cell lines, corpus callosum, cardiac (atria and ventricle), caudate nucleus, CNS and peripheral tissue, cerebellum, cerebral cortex, colon, cortical neurogenic cells, endothelial (coronary artery and umbilical vein) cells, palate epithelia, eye, neonatal eye, frontal cortex, fetal hematopoietic cells, heart, hippocampus, hypothalamus, leukocytes, liver, fetal liver, lung, lung lymphoma cell lines, fetal lymphoid tissue, adult lymphoid tissue, Those that express MHC II and III nervous, medulla, subthalamic nucleus, ovary, pancreas, pituitary, placenta, pons, prostate, putamen, serum, skeletal muscle, small intestine, smooth muscle (coronary artery in aortic) spinal cord, spleen, stomach, taste receptor cells of the tongue, testis, thalamus, and thymus tissue. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV84b polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 84E.

Table 84E. BLAST results for NOV84b

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17474603 ref XP_062552.1 (XM_062552)	similar to olfactory receptor [Homo sapiens]	312	312/312 (100%)	312/312 (100%)	e-145
gi 18479402 gb AAL60715.1 (AY073052)	olfactory receptor MOR160-1 [Mus musculus]	309	216/304 (71%)	254/304 (83%)	e-105
gi 18480580 gb AAL61304.1 (AY073641)	olfactory receptor MOR160-2 [Mus musculus]	308	212/307 (69%)	256/307 (83%)	e-95
gi 18480924 gb AAL61476.1 (AY073813)	olfactory receptor MOR160-5 [Mus musculus]	311	211/309 (68%)	258/309 (83%)	e-94
gi 18480922 gb AAL61475.1 (AY073812)	olfactory receptor MOR160-4 [Mus musculus]	305	184/302 (60%)	233/302 (76%)	6e-84

Table 84F lists the domain descriptions from DOMAIN analysis results against NOV84b. This indicates that the NOV84b sequence has properties similar to those of other proteins known to contain this domain.

5

Table 84F. Domain Analysis of NOV84b

[gnl|Pfam|pfam00001](#), 7tm_1, 7 transmembrane receptor (rhodopsin family). CD-Length = 254 residues, 44.9% aligned
Score = 68.6 bits (166), Expect = 5e-13

The disclosed NOV84b nucleic acid of the invention encoding a GPCR-like protein includes the nucleic acid whose sequence is provided in Table 84A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 84A while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be

used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 38 percent of the bases may be so changed.

5 The disclosed NOV84b protein of the invention includes the GPCR-like protein whose sequence is provided in Table B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 84B while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 50 percent of the residues may be so changed.

10 The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this GPCR-like protein (NOV84b) may function as a member of a “GPCR family”. Therefore, the NOV84b nucleic acids and proteins identified here may be useful in potential therapeutic applications
15 implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues
20 and cell types composing (but not limited to) those defined here.

The NOV84b nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the GPCR-like protein (NOV84b) may be useful in gene therapy, and the GPCR-like protein (NOV84b) may be
25 useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from CNS disorders, brain disorders including epilepsy, eating disorders, schizophrenia, ADD; cancer; heart disease; inflammation and autoimmune disorders including Crohn’s disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, blood disorders;
30 psoriasis colon cancer, leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, pancreatic disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer, or other pathologies or conditions. The NOV84b nucleic acid encoding the GPCR-like protein of the invention, or fragments thereof, may further be useful in diagnostic

NOV85

A disclosed NOV85 nucleic acid of 963 nucleotides (also referred to as CG57839-01) encoding a GPCR-like protein is shown in Table 85A. The start and stop codons are in bold letters.

Table 85A. NOV85 nucleotide sequence (SEQ ID NO:201).

AACATGGAAAGCAATCAGACCTGGATCACAGAAGTCATCCTGTTGGGATTCCAGGTTGGACCAGCTCTGG
AGTTGTTCCCTCTTTGGGTTTTCTTGCTATTCTACAGCTTAACCCTGATGGGAAATTTGGACTCTAGACT
GCACACACCCATGTATGTCTTCTGTGCACATCTGGCCATTGTGGACATGTCCTATGCCTCGAGTACTGTC
CCTAAGATGCTAGCAAATCTTGTGATGCACAAAAAGTCATCTCCTTTGCTCCTTGCATACCTCAGACTT
TTTTGTATTTGGCGTTTGCTATTACAGAGTGTCTGATTTTGGTGATGATGTGCTATGATCGGTATGTGGC
AATCTGTCAACCCCTTGCAATACACCCCTCATTATGAAGTGGAGAGTGTGCACCTGCTGGCCTCAACTGTC
TGGATATTTAGCTTTCTCTTGGCTCTGGTCCATATTACTCTTATTCTGAGGCTGCCTTTTTTGTGGCCACA
AAAGATCAACCACTTTTTTTTTTGTGGCCACAAAAGATCAACCACTTTTTCTGTCAAATCATGTCCGTATT
CAAATTGGCCTGTGCTGACACTAGGCTCAACCAGGTGGTCCTATTGCGGGTTCTGCGTTCATCTTAGTG
GGGCCGCTCTGCCTGGTGTCTGCTCTCTACTTGCACATCCTGGTGGCCATCTTGAGGATCCAGTCTGGGG
AGGGCCGAGAAAGGCCTTCTACCTGCTCCTCCCACCTCTGCGTGGTGGGGCTTTTCTTTGGCAGCGC
CATTGTCATGTACATGGCCCCCAAGTCAAGCCATTCTCAAGAACGGAGGAAGATCCTTTCCCTGTTTTAC
AGCCTTTTCAACCCGATCCTGAACCCCTCATCTACAGCCTTAATGCAGAGGTGAAAGGGGCTCTAAAGA
GAGTCCTTTGGAACAGAGATCAATTGAAGAATCATTTGAGATTTCCTGAGAA

The disclosed NOV85 polypeptide (SEQ ID NO:202) encoded by SEQ ID NO:201 has 318 amino acid residues and is presented in Table 85B using the one-letter amino acid code

Table 85B. Encoded NOV85 protein sequence (SEQ ID NO:202).

MESNQTWITEVILLGFQVGPALFLFLFGFFLLFYSLTLMGNLDSRLHTPMYVFLSHLAIVDMSYASSTVP
KMLANLVMHKKVISFAPCILQTFLYLAFATECLILVMCYDRYVAICHPLQYTLIMNWRVCTVLASTCW
IFSFLALVHITLILRLPFCGHKRSTTFFLWPQKINHFFCQIMSVFKLACADTRLNQVVLFAFSAFILVG
PLCLVLVSYLHILVAILRIQSGEGRRKAFSTCSSHLCVVGLFFGSAIVMYMAPKSSHSQERRKILSLFYS
LFNPILNPLIYSLNAEVKGALKRVLWKQRSIEESFEIS

A search of sequence databases reveals that the NOV85 amino acid sequence has 181/311 (58%) identity and 232/311 (74%) similarity with SPTREMBL-ACC:O95047 WUGSC:H_DJ0988G15.2 PROTEIN - Homo sapiens. Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

- 5 The disclosed NOV85 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 85C.

Table 85C. BLAST results for NOV85					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 18565914 ref XP_095005.1 (XM_095005)	protein XP_095005 [Homo sapiens]	310	299/320 (93%)	300/320 (93%)	e-134
gi 18480894 gb AAL61461.1 (AY073798)	olfactory receptor MOR261- 12 [Mus musculus]	308	254/317 (80%)	277/317 (87%)	e-115
gi 18480182 gb AAL61105.1 (AY073442)	olfactory receptor MOR261-4 [Mus musculus]	310	238/320 (74%)	266/320 (82%)	e-111
gi 18480180 gb AAL61104.1 (AY073441)	olfactory receptor MOR261-3 [Mus musculus]	310	235/320 (73%)	263/320 (81%)	e-110
gi 18565912 ref XP_069619.2 (XM_069619)	similar to olfactory receptor [Homo sapiens]	311	231/320 (72%)	258/320 (80%)	e-108

- 10 Table 85D lists the domain descriptions from DOMAIN analysis results against NOV85. This indicates that the NOV85 sequence has properties similar to those of other proteins known to contain this domain.

Table 85D. Domain Analysis of NOV85
gnl Pfam pfam00001 , 7tm_1, 7 transmembrane receptor (rhodopsin family).
CD-Length = 254 residues, 94.9% aligned
Score = 111 bits (278), Expect = 5e-26

- 15 The disclosed NOV85 nucleic acid of the invention encoding a GPCR-like protein includes the nucleic acid whose sequence is provided in Table 85A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 85A while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a fragment of such a nucleic acid. The

invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 5 percent of the bases may be so changed.

The disclosed NOV85 protein of the invention includes the GPCR-like protein whose sequence is provided in Table 85B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 85B while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 42 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this GPCR-like protein (NOV85) may function as a member of a “GPCR family”. Therefore, the NOV85 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV85 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the GPCR-like protein (NOV85) may be useful in gene therapy, and the GPCR-like protein (NOV85) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from CNS disorders, brain disorders including epilepsy, eating disorders, schizophrenia, ADD;

The disclosed NOV86 polypeptide (SEQ ID NO:204) encoded by SEQ ID NO:203 has 319 amino acid residues and is presented in Table 86B using the one-letter amino acid code.

Table 86B. Encoded NOV86 protein sequence (SEQ ID NO:204).
MAFGEFSPPIPPSSLLPLGLFCGPPKQSRGFLPFFVLFLGIYLLTIMENLMLLLMIRADSLHKPMYFFL SHLSFVDLCFSSVIVPKMLENLLSQRTISVEGCLAQVFFVFTAGTEACLLSGMAYDRHAAICRPLLYG QIMGKQLYMHVLVWGSWGLGFLDALINVLAVNMVFCEAKIHHYSYEMPSLLPLSCSDISRSLIALLCST LLHGLGNFLVFLSYTRIISTILSISSTSGRSKAFSTCSAHLTAVTLYYGSGLLRHLMPSGSPIELIFS VQYTVVTPMLNSLIYSLKNKEVKVALKRTLEKYLYTRR

A search of sequence databases reveals that the NOV86 amino acid sequence has 134/300 (44%) identity and 192/300 (64%) similarity with SPTREMBL-ACC:O35184 OLFACTORY RECEPTOR - Rattus norvegicus. Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

The disclosed NOV86 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 86C.

Table 86C. BLAST results for NOV86					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17474603 ref XP_062552.1 (XM_062552)	similar to olfactory receptor [Homo sapiens]	312	290/307 (94%)	293/307 (94%),	e-134
gi 18479402 gb AAL60715.1 (AY073052)	olfactory receptor MOR160-1 [Mus musculus]	309	200/281 (71%)	234/281 (83%)	3e-97
gi 18480580 gb AAL61304.1 (AY073641)	olfactory receptor MOR160-2 [Mus musculus]	308	193/280 (68%)	236/280 (83%)	8e-88
gi 18480924 gb AAL61476.1 (AY073813)	olfactory receptor MOR160-5 [Mus musculus]	311	192/282 (68%)	237/282 (83%)	5e-87
gi 18480922 gb AAL61475.1 (AY073812)	olfactory receptor MOR160-4 [Mus musculus]	305	170/279 (60%)	215/279 (76%)	e-78

Table 86D lists the domain descriptions from DOMAIN analysis results against NOV86. This indicates that the NOV86 sequence has properties similar to those of other proteins known to contain this domain.

Table 86D. Domain Analysis of NOV86

gnl|Pfam|pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family).

CD-Length = 254 residues, 44.9% aligned

Score = 68.2 bits (165), Expect = 7e-13

The disclosed NOV86 nucleic acid of the invention encoding a GPCR-like protein includes the nucleic acid whose sequence is provided in Table 86A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed
5 from the corresponding base shown in Table 86A while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or
10 complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.
15 In the mutant or variant nucleic acids, and their complements, up to about 5 percent of the bases may be so changed.

The disclosed NOV86 protein of the invention includes the GPCR-like protein whose sequence is provided in Table 86B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 86B
20 while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 56 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

25 The above defined information for this invention suggests that this GPCR-like protein (NOV86) may function as a member of a "GPCR family". Therefore, the NOV86 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein

therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

5 The NOV86 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the GPCR-like protein (NOV86) may be useful in gene therapy, and the GPCR-like protein (NOV86) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the
10 compositions of the present invention will have efficacy for treatment of patients suffering from CNS disorders, brain disorders including epilepsy, eating disorders, schizophrenia, ADD; cancer; heart disease; inflammation and autoimmune disorders including Crohn's disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, blood disorders; psoriasis colon cancer, leukemia AIDS; thalamus disorders; metabolic disorders including
15 diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, pancreatic disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer, or other pathologies or conditions. The NOV86 nucleic acid encoding the GPCR-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be
20 assessed.

 NOV86 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV86 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-
25 NOVX Antibodies" section below. The disclosed NOV86 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

30 **NOV87**

 A disclosed NOV87 nucleic acid of 1067 nucleotides (also referred to as CG56763-01) encoding a GPCR-like protein is shown in Table 87A. The start and stop codons are in bold letters.

Table 87A. NOV87 nucleotide sequence (SEQ ID NO:205).

CCCTTTCCTCTTGCTCTTTGATGTTTTGTAGGCCTGCAGCTCCCAAGCACAGAGGCATGAGTGGGGAGAA
TGTCACCAAGGGTCGGCACCTTCATCCTGGTGGGCTTCCCCACGGCCCCAGGGCTGCAGTACCTGCTCTTC
CTCCTCTTCTGCTCACCTACCTCTTTGTCTGGTGGAGAACCCTGGCCATCATCCTCACCCTGCGGAGCA
GCACCTCCCTCCACAGGCCCATGTACTACTTTCTGAGCTCCATGTCTTTCTAGAGATCTGGTACGTGTC
TGACATCACCCCAAGATGCTGGAGGGCTTCTCCTCCAGCAGAAACGCATCTCTTTCTGTCGGGTGCATG
ACGCAGCTCTACTTCTCAGCTCCCTGGTGTGCACCGAGTGTGTGCTTCTGGCCTCCATGGCCTACGACC
GCTACGTGGCCATCTGCCACCCGCTGCGCTACCACGTCCTTGTGACCCCGGGGCTGTGCCTCCAGCTGGT
GGGCTTCTCCTTTGTGAGTGGCTTACCATCTCCATGATCAAGGTCTGTTTTATCTCCAGCGTCACGTTT
TGTGGCTCCAACGCTCTGAACCACTTCTTCTGTGACATTTCCCCATCCTCAAGCTGGCCTGCACGGACT
TCTCCACTGCAGAGCTGGTGGATTTCATTCTGGCCTTCATCATCCTGGTGTTCCTACTCTGGCCACCAT
GCTGTATATGCGACATCACCTGGCTGTCTGCGCATCCCTCGGCCACCGGTGCTGGAGAGCCTTC
TTCACCTGCGCCTCTACCTCACCGTGGTCACCGTCTTCTATACAGCCTTGCTTTTCATGTATGTCCGGC
CCCAGGCCATTGATTCCCGGAGCTCCAACAAGCTCATCTCTGTTTTGTACACAGTTATACCCCCATCTT
GAACCCCTTGATATACTGCTGAGGAATAAGGAATTTAAGAATGCCTTGAAAAAGCCTTCGGCTTGACG
AGCTGCGCCGTAGAGGGGAGGCTTCTAGTCTTCTGGAACCTTCATCTCCAAATACACAGCCAGCCTCTCT
GAGGAGGCCATTTGACT

The disclosed NOV87 polypeptide (SEQ ID NO:206) encoded by SEQ ID NO:205
has 343 amino acid residues and is presented in Table 87B using the one-letter amino acid
code.

Table 87B. Encoded NOV87 protein sequence (SEQ ID NO:206).

MFCRPAAPKHRGMSGENVTRVGTFILVGFPTAPGLQYLLFLLFLLTYLFVLVENLAIILTVWSSTSLHRP
MYYFLSSMSFLEIWVSDITPKMLEGFLQKRIISFVGCMTQLYFFSSSLVCTECVLLASMAVDYVAICH
PLRYHVLVTPGLCLQLVGFSFVSGFTISMIKVCFISSVTFCGSNVLNHHFCDISPILKLACTDFSTAEV
DFILAFIILVFPLLATMLSYAHITLAVLRIPSATGCWRAFFTCASHLTVVTVFYTALLFMYVRPQAIDSR
SSNKLISVLYTVITPILNPLIYCLRNKEFKNALKKAFLTSCAVERGLSSLELHLQIHSQPL

A search of sequence databases reveals that the NOV87 amino acid sequence has
211/306 (68%) identity and 250/306 (81%) similarity with TREMBLNEW-ACC:AAG45189
M51 OLFACTORY RECEPTOR - Mus musculus. Public amino acid databases include the
GenBank databases, SwissProt, PDB and PIR.

The disclosed NOV87 polypeptide has homology to the amino acid sequences shown
in the BLASTP data listed in Table 87C.

Table 87C. BLAST results for NOV87

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17435880 ref XP 065375.1 (XM_065375)	similar to olfactory receptor 41 [Homo sapiens]	331	331/331 (100%)	331/331 (100%)	e-150
gi 17435888 ref XP 065377.1 (XM_065377)	similar to olfactory receptor 41 [Homo sapiens]	312	293/307 (95%)	300/307 (97%)	e-133

<u>gi 18479396 gb AAL60712.1 </u> (AY073049)	olfactory receptor MOR103-2 [Mus musculus]	312	270/307 (87%)	287/307 (92%)	e-122
<u>gi 18479398 gb AAL60713.1 </u> (AY073050)	olfactory receptor MOR103-3 [Mus musculus]	312	261/307 (85%)	284/307 (92%)	e-120
<u>gi 12007416 gb AAG45189.1 </u> (AF321234)	m51 olfactory receptor [Mus musculus]	314	211/306 (68%)	250/306 (80%)	e-101

Table 87D lists the domain descriptions from DOMAIN analysis results against NOV87. This indicates that the NOV87 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 87D. Domain Analysis of NOV87	
<u>gnl Pfam pfam00001</u> , 7tm_1, 7 transmembrane receptor (rhodopsin family).	
CD-Length = 254 residues, 100.0% aligned	
Score = 98.2 bits (243), Expect = 7e-22	

The disclosed NOV87 nucleic acid of the invention encoding a GPCR-like protein includes the nucleic acid whose sequence is provided in Table 87A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 87A while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 5 percent of the bases may be so changed.

The disclosed NOV87 protein of the invention includes the GPCR-like protein whose sequence is provided in Table 87B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 87B while still encoding a protein that maintains its GPCR-like activities and physiological

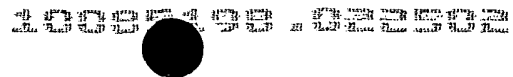
functions, or a functional fragment thereof. In the mutant or variant protein, up to about 32 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

5 The above defined information for this invention suggests that this GPCR-like protein (NOV87) may function as a member of a “GPCR family”. Therefore, the NOV87 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein
10 therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV87 nucleic acids and proteins of the invention are useful in potential
15 therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the GPCR-like protein (NOV87) may be useful in gene therapy, and the GPCR-like protein (NOV87) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering
20 from CNS disorders, brain disorders including epilepsy, eating disorders, schizophrenia, ADD; cancer; heart disease; inflammation and autoimmune disorders including Crohn’s disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, blood disorders; psoriasis colon cancer, leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, pancreatic
25 disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer, or other pathologies or conditions. The NOV87 nucleic acid encoding the GPCR-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

30 NOV87 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV87 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV87 proteins have multiple hydrophilic



regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

5 NOV88

A disclosed NOV88 nucleic acid of 939 nucleotides (also referred to as CG56753-01) encoding a GPCR-like protein is shown in Table 88A. The start and stop codons are in bold letters.

Table 88A. NOV88 nucleotide sequence (SEQ ID NO:207).

ATGTCAGGAGAAAATAATTCCTCAGTGACTGAGTTCATTCTGGCTGGGCTCTCAGAACAGCCAGAGCTCC
AGCTGCCCCCTCTTCCTCCTGTTCTTAGGAATCTATGTGGTCACAGTGGTGGGCAACCTGGGCATGACCAC
ACTGATTGGCTCAGTTCCTCACCTGCACACCCCTATGTACTATTTCCTCAGCAGTCTGTCTTCATTGAC
TTCTGCCATTCCACTGTCATTACCCCTAAGATGCTGGTGAACCTTGTGACAGAGAAGAACATCATCTCCT
ACCCTGAATGCATGACTCAGCTCTACTTCTTCCTCGTTTTTGTCTATTGCAGAGTGTACATGTTGGCTGC
AATGGCGTATGACCGTTACATGGCCATCTGTAGCCCTTGCTGTACAGTGTATCATATCCAATAAGGCT
TGCTTTCTCTGATTTAGGGGTGTATATAATAGGCCTGGTTTGTGCATCAGTTCATACAGGCTGTATGT
TTAGGGTCAATTCTGCAAATTTGATTTGATTAAACCATTATTTCTGTGATCTCTTCCCTCCTAAAGCT
CTCTTGCTCTAGTATCTATGTCAACAACTACTTATTCTATGTGTTGGTGCATTTAACATCCTTGTCCCC
AGCCTGACCATCCTTTGCTCTTACATCTTTATTATTGCCAGCATCCTCCACATTCGCTCCACTGAGGGCA
GGTCCAAAGCCTTCAGCACTTGTAGCTCCACATGTTGGCGGTGTAAATCTTTTTGGATCTGCAGCATT
CATGTACTTGCAGCCATCTTCAATCAGCTCCATGGACCAGGGGAAAGTATCCTCTGTGTTTTATACTATT
ATTGTGCCCATGTTGAACCTCTGATTTATAGCCTGAGGAATAAAGATGTCCATGTTTCCCTGAAGAAA
TGCTACAGAGAAGAACATTATTG**TAAACA**

The disclosed NOV88 polypeptide (SEQ ID NO:208) encoded by SEQ ID NO:207 has 311 amino acid residues and is presented in Table 88B using the one-letter amino acid code.

Table 88B. Encoded NOV88 protein sequence (SEQ ID NO:208).

MSGENNSSVTEFILAGLSEQPELQLPLFLLFLGIYVVTVVGNLGMTTLIWLSSHLHTPMYYFLSSLSFID
FCHSTVITPKMLVNFVTEKNIISYPECMTQLYFFLVFAIAECHMLAAMAYDRYMAICSPLLYSVIIISNKA
CFSLILGVYIIIGLVCASVHTGCMFRVQFCKFDLINHYFCDLLPLLKLSCSSIYVNKLLILCVGAFNII
SLTILCSYIFIIASILHIRSTEGRSKAFSTCSSHMLAVIIFGSAAFMYLQPSISSMDQGVSSVFYTI
IVPMLNPLIYSLRNKDVHVSLLKMLQRRTLL

A search of sequence databases reveals that the NOV88 amino acid sequence has 239/311 (76%) identity and 275/311 (88%) similarity with TREMBLNEW-ACC:AAG39856 ODORANT RECEPTOR K11 - Mus musculus. Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

The disclosed NOV88 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 88C.

Table 88C. BLAST results for NOV88					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
<u>gi 18578547 ref XP_090109.1 </u> (XM_090109)	olfactory receptor, family 8, subfamily G, member 1 [Homo sapiens]	311	311/311 (100%)	311/311 (100%)	e-141
<u>gi 17472672 ref XP_061794.1 </u> (XM_061794)	similar to odorant receptor K11 [Homo sapiens]	311	263/311 (84%)	285/311 (91%)	e-122
<u>gi 18479824 gb AAL6_0926.1 </u> (AY073263)	olfactory receptor MOR171- 5 [Mus musculus]	314	240/311 (77%)	276/311 (88%)	e-114
<u>gi 11692519 gb AAG3_9856.1 AF282271.1</u> (AF282271)	odorant receptor K11 [Mus musculus]	314	239/311 (76%)	275/311 (87%)	e-113
<u>gi 17472670 ref XP_061793.1 </u> (XM_061793)	similar to odorant receptor K15 [Homo sapiens]	258	258/258 (100%)	258/258 (100%)	e-113

Table 88D lists the domain descriptions from DOMAIN analysis results against NOV88. This indicates that the NOV88 sequence has properties similar to those of other proteins known to contain this domain.

Table 88D. Domain Analysis of NOV88	
<u>gnl Pfam pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family).</u>	CD-Length = 254 residues, 100.0% aligned
	Score = 88.6 bits (218), Expect = 5e-19

The disclosed NOV88 nucleic acid of the invention encoding a GPCR-like protein includes the nucleic acid whose sequence is provided in Table 88A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 88A while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be

used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 5 percent of the bases may be so changed.

5 The disclosed NOV88 protein of the invention includes the GPCR-like protein whose sequence is provided in Table 88B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 88B while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 24 percent of the residues may be so changed.

10 The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this GPCR-like protein (NOV88) may function as a member of a “GPCR family”. Therefore, the NOV88 nucleic acids and proteins identified here may be useful in potential therapeutic applications
15 implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues
20 and cell types composing (but not limited to) those defined here.

The NOV88 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the GPCR-like protein (NOV88) may be useful in gene therapy, and the GPCR-like protein (NOV88) may be useful
25 when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from CNS disorders, brain disorders including epilepsy, eating disorders, schizophrenia, ADD; cancer; heart disease; inflammation and autoimmune disorders including Crohn’s disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, blood disorders;
30 psoriasis colon cancer, leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, pancreatic disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer, or other pathologies or conditions. The NOV88 nucleic acid encoding the GPCR-like protein of the invention, or fragments thereof, may further be useful in diagnostic

NOV89

A disclosed NOV89 nucleic acid of 1003 nucleotides (also referred to as CG57670-01) encoding a GPCR-like protein is shown in Table 89A. The start and stop codons are in bold letters.

Table 89A. NOV89 nucleotide sequence (SEQ ID NO:209).

TGTTCCATACATTATTTTGTCTTTTGTCTGAAGCAATGCTGAATACAACCTCAGTCACTGAATT
TCTCCTTTTGGGAGTGACAGACATTCAAGAACTGCAGCCTTTTCTCTTCGTTGTTTTCCTTACC
ATCTACTTTCATCAGTGTGGCTGGGAATGGAGCCATTCTGATGATTGTCATCTCTGATCCTAGAC
TCCATTCCCCTATGTATTTCTTCTGGGAAACCTGTCTCTGCACACAAAGCAATTTCTTCTTGGGATGC
AACACTGCCAAAAATGCTGCAGAACTTCTCTCTGCACACAAAGCAATTTCTTCTTGGGATGC
ATAAGCCAACCTCCATTTCTTCCACTTCTTGGGAGCAGAGGCCATGTTGTTGGCCGTGATGG
CATTTGACCGCTTTGTGGCTATTTGCAAGCCACTTCGCTACACTGTCATTATGAACCCCTCAGCT
CTGTACCCAGATGGCCATCACAACTCTGGATGATTGGTTTTTCCATGCCCTGCTGCACTCCCTA
ATGACCTCTCGCTTGAACCTTCTGTGGTTCTAACCGTATCTATCACTTCTTCTGTGATGTGAAGC
CATTGCTAAAGCTGAGCCTTAATCAGTGGCTGCTCAGTACTGTACAGGGACAATCGCCATGGG
CCCCTTCTTTCTCACATTACTCTCCTATTTCTACATTATCACCCATCTCTTCTTCAAGACTCAT
TCTTTTAGCATGCTCCGAAAGCACTGTCCACTTGTGCCTCCCACTTCATGGTAGTTATTCTTT
TGTATGCACCTGTTCTTTCACCTATATTATCATATGCCTCAGGGACCTCCATGGACCAGGACCG
GATCACTGCCATCATGTATACTGTGGTCACTCCAGTACTAAACCCACTGATCTACACTTTGAGG
AACAAGGAAGTGAAAGGGGCTTTAATAGAGCAATGAAAAGGTGGCTTTGGCCTAAAGAAATCT
TGAAGAAGCTCTTCTGAAGCATAAATAACAATTAAAAAGATGA

The disclosed NOV89 polypeptide (SEQ ID NO:210) encoded by SEQ ID NO:209 has 315 amino acid residues and is presented in Table 89B using the one-letter amino acid code

Table 89B. Encoded NOV89 protein sequence (SEQ ID NO:210).

MLNTTSVTEFLLLVGTDIQLQPFLLVFLTIYFISVANGAILMIVISDPRHLHSPMYFFLGNLSCLDIC
YSSVTLPKMLQNFSLAHKAISFLGCISQLHFFHFLGSTEAMLLAVMAFDRFVAICKPLRYTVIMNPQLCT
QMAITIMIGFFHALLHSLMTSRLNFCGSNRIYHFFCDVKPLKLSLNQWLLSTVTGTIAMGPFFLTLS
YFYIITHLFFKTHSFSMLRKALSTCASHFMVVILLYAPVLFTYIHHASGTSMDQDRITAIMYTVVTPVLN

PLIYTLRNKEVKGAFNRAMKRWLWPKEILKNSSEA

A search of sequence databases reveals that the NOV89 amino acid sequence has similarity with SPTREMBL-ACC:Q9UGF7 BA150A6.1 (NOVEL 7 TRANSMEMBRANE RECEPTOR (RHODOPSIN FAMILY)(OLFACTORY RECEPTOR LIKE) PROTEIN (HS6M1-27)) - Homo sapiens. Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

The disclosed NOV89 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 89C.

Table 89C. BLAST results for NOV89					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17464355 ref XP_069464.1 (XM_069464)	similar to olfactory receptor, family 12, subfamily D, member 2 [Homo sapiens]	284	279/315 (88%)	280/315 (88%)	e-130
gi 18563691 ref XP_094753.1 (XM_094753)	hypothetical protein XP_094753 [Homo sapiens]	284	278/315 (88%)	279/315 (88%)	e-130
gi 7363443 ref NP_039224.1 (NM_013936)	olfactory receptor, family 12, subfamily D, member 2 [Homo sapiens]	307	269/306 (87%)	281/306 (90%)	e-126
gi 18563689 ref XP_084201.1 (XM_084201)	similar to olfactory receptor, family 12, subfamily D, member 2 (H. sapiens) [Homo sapiens]	307	268/306 (87%)	280/306 (90%)	e-125
gi 15020328 emb CAC44545.1 (AL133159)	bM332P19.2 (novel 7 transmembrane receptor (rhodopsin family) (olfactory receptor like) protein (mm17M1- 13); ortholog of human DJ994E9.8 (HS6M1-20)) [Mus musculus]	308	241/308 (78%)	269/308 (87%)	e-117

Table 89D lists the domain descriptions from DOMAIN analysis results against NOV89. This indicates that the NOV89 sequence has properties similar to those of other proteins known to contain this domain.

Table 89D. Domain Analysis of NOV89

gnl|Pfam|pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family).

CD-Length = 254 residues, 100.0% aligned

Score = 101 bits (251), Expect = 7e-23

The disclosed NOV89 nucleic acid of the invention encoding a GPCR-like protein includes the nucleic acid whose sequence is provided in Table 89A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed
5 from the corresponding base shown in Table 89A while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or
10 complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.
15 In the mutant or variant nucleic acids, and their complements, up to about 5 percent of the bases may be so changed.

The disclosed NOV89 protein of the invention includes the GPCR-like protein whose sequence is provided in Table 89B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 89B
20 while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 5 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

25 The above defined information for this invention suggests that this GPCR-like protein (NOV89) may function as a member of a "GPCR family". Therefore, the NOV89 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein

therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

5 The NOV89 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the GPCR-like protein (NOV89) may be useful in gene therapy, and the GPCR-like protein (NOV89) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the
10 compositions of the present invention will have efficacy for treatment of patients suffering from CNS disorders, brain disorders including epilepsy, eating disorders, schizophrenia, ADD; cancer; heart disease; inflammation and autoimmune disorders including Crohn's disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, blood disorders; psoriasis colon cancer, leukemia AIDS; thalamus disorders; metabolic disorders including
15 diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, pancreatic disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer, or other pathologies or conditions. The NOV89 nucleic acid encoding the GPCR-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be
20 assessed.

NOV89 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV89 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-
25 NOVX Antibodies" section below. The disclosed NOV89 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

30 **NOV90**

A disclosed NOV90 nucleic acid of 950 nucleotides (also referred to as CG57676-01) encoding a GPCR-like protein is shown in Table 90A. The start and stop codons are in bold letters.

Table 90A. NOV90 nucleotide sequence (SEQ ID NO:211).

GTAATAGGAAATGAATGATGATGGAAAAGTCAATGCTAGCTCTGAGGGGTACTTT
 ATTTTAGTTGGATTTTCTAATTGGCCTTATCTGGAAGTAGTTCTCTTTGTGGTTATTTTGATCTTCTGCT
 TGATGACACTGATAGGAAACCTGTTTCATCATCATCCTGACGTACCTGGACTCCCATCTCCATACTCCCTT
 GTATTTCTTCTTTCAAATCTCTCATTTCTGGATCTCTGCTACACCACCAGCTCTATCCCTCAGTTGCTG
 GTCAGTCTCTGGGGTGTGGAAAAGACCATTTCTTATGCTGGTTGCATGGTTCAACTTTACTTTTTTCTCA
 CACTGGGAACACAGAGTGTGCTCTACTGGTGGTATGTCCTATGACCGTTATGCAGCTGTGTGTAGACC
 TTTGCATTACACTGTCTCATGCACTCTCGTTTCTGCCACTTGTGGCTGTGGCTTCTTGGGTAAGTGGT
 TTTACAAACCCAGCACTTCATTCCTCCTTCACCTTCTGGGTACCTCTGTGTGGACACCGCCAAATAGATC
 ACTTTTTCTGTGAAGTTCCGGCACTTTTAAGATTATCATTTGTCAATACCGTGAAAATAAACTGACCCT
 CATGATCACAGCTCCATTTTGTCTGCTACTTCTCACCTCATTTTCACTTCTATGGTGTCTATTGCC
 CAGGCTGTACTGAGGATGCAGTCAACCACTGGGCTTCAGAAAGTATTTGGAACATGTGGAGCTCATCATA
 TGGTTGTATCTCTTTTTTCATTCCGGCCATGTGCATGTATCTCCAGCCACCATCAGGGAATTCTCAAGA
 TCAAGGCAAGTTCATTGCTCTCTTTTATACTGTTGTACACCTAGTCTTAACCCTCTAATCTACACCCTC
 AGAAACAAAGATGTAAGAGGGGTAGTGAAGAGACTAAGGGGTGGGAGTGAGCCT

The disclosed NOV90 polypeptide (SEQ ID NO:212) encoded by SEQ ID NO:211 has 311 amino acid residues and is presented in Table 90B using the one-letter amino acid code.

Table 90B. Encoded NOV90 protein sequence (SEQ ID NO:212).

MNDDGKVNASSEGYFILVGFSNWPYLEVVLFFVILIFCLMTLIGNLFIILTYLDSHLHTPLYFFLSNLS
 FLDLCYTTSSIPQLLVSLWGVEKTSYAGCMVQLYFFLTGLTTECVLLVMSYDRYAACRPLHYTVLMH
 SRFCHLLAVASWVSGFTNPALHSSFTFWVPLCGHRQIDHFFCEVPALLRLSFVNTRENKLTLMITSSIFV
 LLLLTLLIFTSYGAIAQAVLRMQSTTGLQKVFGTCGAHMMVSLFFIPAMCMYLQPPSGNSQDQGFIALF
 YTVVTPSLNPLIYTLRNKDVGRGVVKRLRGWE

A search of sequence databases reveals that the NOV90 amino acid sequence has similarity with TREMBLNEW-ACC:CAC20478 OLFACTORY RECEPTOR - Homo sapiens. Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

The disclosed NOV90 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 90C.

Table 90C. BLAST results for NOV90

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 12054347 emb CAC20478.1 (AJ302558)	olfactory receptor [Homo sapiens]	311	281/311 (90%)	293/311 (93%)	e-139
gi 12054345 emb CAC20477.1 (AJ302557)	olfactory receptor [Homo sapiens]	311	280/311 (90%)	293/311 (94%)	e-139
gi 12140469 emb CAC21440.1 (AJ302552)	olfactory receptor [Homo sapiens]	311	280/311 (90%)	292/311 (93%)	e-139

gi 14423775 sp O760 01 O2J3 HUMAN	OLFACTORY RECEPTOR 2J3 (OLFACTORY RECEPTOR 6-6) (OR6-6)	311	280/311 (90%)	293/311 (94%)	e-139
gi 18564769 ref XP 069457.2 (XM_069457	similar to OLFACTORY RECEPTOR 2J3 (OLFACTORY RECEPTOR 6-6) (OR6-6) (HS6M1-3) [Homo sapiens]	391	270/298 (90%)	283/298 (94%)	e-136

Table 90D lists the domain descriptions from DOMAIN analysis results against NOV90. This indicates that the NOV90 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 90D. Domain Analysis of NOV90	
gnl Pfam pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family).	CD-Length = 254 residues, 100.0% aligned
	Score = 89.7 bits (221), Expect = 2e-19

The disclosed NOV90 nucleic acid of the invention encoding a GPCR-like protein includes the nucleic acid whose sequence is provided in Table 90A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 90A while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 5 percent of the bases may be so changed.

The disclosed NOV90 protein of the invention includes the GPCR-like protein whose sequence is provided in Table 90B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 90B

while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 5 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this GPCR-like protein (NOV90) may function as a member of a "GPCR family". Therefore, the NOV90 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV90 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the GPCR-like protein (NOV90) may be useful in gene therapy, and the GPCR-like protein (NOV90) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from CNS disorders, brain disorders including epilepsy, eating disorders, schizophrenia, ADD; cancer; heart disease; inflammation and autoimmune disorders including Crohn's disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, blood disorders; psoriasis colon cancer, leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, pancreatic disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer, or other pathologies or conditions. The NOV nucleic acid encoding the GPCR-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV90 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV90 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-

NOVX Antibodies” section below. The disclosed NOV90 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV91

A disclosed NOV91 nucleic acid of 967 nucleotides (also referred to as CG57668-01) encoding a GPCR-like protein is shown in Table 91A. The start and stop codons are in bold letters.

Table 91A. NOV91 nucleotide sequence (SEQ ID NO:213).

GCATTTGCCCCAGTAGCTATGATTATAATTTGCAATGACAGCCACAGTGATTTTCATCCTTCTGG
GCTTCTCTAACAGCCACATTTGGAGAAGATACTTTTTTGGATCATTTTTTATTTTTTATTTTTT
GACTCTTGCAGGAAATATGGTCATAGTTCTTGTGTCCTTGAAGGATCCAAAACCTCCACATCCCT
ATGTATTTCTTCTTCCAAACCTTTCCTTGGTAGACCTCTGTTTGACCAGCAGCTGTGTTCCAC
AGATGTTGATTAACTTCTGGGGCCAGAAAAGACCATCAGCTACATTGGCTGTGCCATTCAACT
CTATGTTTTTTTGTGGCTTGGGGCCACGGAATATGTCCTTCTTGTGTCATGGCTGTGGATTGT
TATGTAGCAGTGTGTCATCCACTGCAAAATACCATGATCATGCACCCAAAACCTTGTCTGCAGC
TGGCTATCTTGGCATGGGGGACTGGCTTGGCCAGTCTCTGATCCAGTCCCCTGCCACCCTCCG
GTTACCCCTTCTGCTCCCAGCGGATGGTGGATGATGTTGTTTGTGAAGTCCCAGCTCTGATTCAG
CTCTCCAGTACTGATACTACCTACAGTGAATTCAGATGTCTATCGCCAGTGTGTCTCTCTGG
TGATGCCCTTGATCATATCCTTTCCTCTTCTGGTGCTATTGCTAAGGCTGTGCTGAGAATTAA
GTCAACTGCAGGACAGAAGAAAGCATTGGGCACCTGCATCTCTACCTTCTTGTGGTTTCTCTC
TTTTATGGCACTGTACAGGTGTCTACCTTCAACCAAAAATCACTATCCTCATGAATGGGGCA
AATTTCTCACTCTTTTCTACACTGTAGTAACCCCAACTCTTAATCCCTCATCTACACTCTAAG
GAACAAGGAGGTAAAGGGAGCACTAATAAGATTGGGGAGGAGGACCTGGGATTCCCAGAATAAC
TAACAAG

The disclosed NOV91 polypeptide (SEQ ID NO:214) encoded by SEQ ID NO:213 has 314 amino acid residues and is presented in Table 91B using the one-letter amino acid code.

Table 91B. Encoded NOV91 protein sequence (SEQ ID NO:214).

MIIICNDSHSDFILLGFSNKPHEKILFWIIFIFYFLTLAGNMVIVLVSLKDPKLHIPMYFFLSNLSLVD
LCLTSSCVPQMLINFWGPEKTISYIGCAIQLYVFLWLGATEYVLLVVMVAVDCYVAVCHPLQNTMIMHPKL
CLQLAILAWGTGLAQSLIQSPATRLRPFCSQRMVDDVCEVPALIQLSSTDTTYSEIQMSIASVLLVMP
LIIILSSSGAIAKAVLRIKSTAGQKKAFTGTCISHLVLSLFYGTVTGVYLPKNHYPHWKGKFLTLFYTV
VTPTLNPLIYTLRNKEVKGALIRLRRTWDSQNN

A search of sequence databases reveals that the NOV91 amino acid sequence has 188/303 (62%) identity and 232/303 (76%) similarity with SWISSNEW-ACC:Q15062 OLFACTORY RECEPTOR 2H3 (OLFACTORY RECEPTOR-LIKE PROTEIN FAT11) -

Homo sapiens. Public amino acid databases include the GenBank databases, SwissProt, PDB and PIR.

The disclosed NOV91 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 91C.

5

Table 91C. BLAST results for NOV91					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17464345 ref XP_069459.1 (XM_069459)	similar to olfactory receptor [Homo sapiens]	470	217/221 (98%)	219/221 (98%)	e-111
gi 18565396 ref XP_094938.1 (XM_094938)	216/221 (97%) , Positives = 218/221 (97%)	289	216/221 (97%)	218/221 (97%)	e-111
gi 12231029 sp Q15062 O2H3 HUMAN	Olfactory receptor 2H3 (Olfactory receptor-like protein FAT11)	316	188/303 (62%)	232/303 (76%)	6e-99
gi 9798922 gb AAF98753.1 AF211941.1 (AF211941)	olfactory receptor [Homo sapiens]	303	187/293 (63%)	228/293 (76%)	2e-97
gi 14423783 sp O95918 O2H2 HUMAN	Olfactory receptor 2H2 (Hs6M1-12)	312	187/302 (61%)	230/302 (75%)	3e-97

Table 91D lists the domain descriptions from DOMAIN analysis results against NOV91. This indicates that the NOV91 sequence has properties similar to those of other proteins known to contain this domain.

10

Table 91D. Domain Analysis of NOV91	
gnl Pfam pfam00001 , 7tm_1, 7 transmembrane receptor (rhodopsin family).	
CD-Length = 254 residues, 100.0% aligned	
Score = 98.6 bits (244), Expect = 5e-22	

The disclosed NOV91 nucleic acid of the invention encoding a GPCR-like protein includes the nucleic acid whose sequence is provided in Table 91A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 91A while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just

15

described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 5 percent of the bases may be so changed.

The disclosed NOV91 protein of the invention includes the GPCR-like protein whose sequence is provided in Table 91B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 91B while still encoding a protein that maintains its GPCR-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 38 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above defined information for this invention suggests that this GPCR-like protein (NOV91) may function as a member of a "GPCR family". Therefore, the NOV91 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV91 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diseases including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the GPCR-like protein (NOV91) may be useful in gene therapy, and the GPCR-like protein (NOV91) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from CNS disorders, brain disorders including epilepsy, eating disorders, schizophrenia, ADD; cancer; heart disease; inflammation and autoimmune disorders including Crohn's disease,

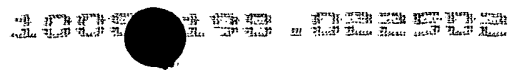
IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, blood disorders; psoriasis colon cancer, leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, pancreatic disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer, or other pathologies or conditions. The NOV nucleic acid encoding the GPCR-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV91 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV91 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV91 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOVX Nucleic Acids and Polypeptides

One aspect of the invention pertains to isolated nucleic acid molecules that encode NOVX polypeptides or biologically active portions thereof. Also included in the invention are nucleic acid fragments sufficient for use as hybridization probes to identify NOVX-encoding nucleic acids (*e.g.*, NOVX mRNAs) and fragments for use as PCR primers for the amplification and/or mutation of NOVX nucleic acid molecules. As used herein, the term “nucleic acid molecule” is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments and homologs thereof. The nucleic acid molecule may be single-stranded or double-stranded, but preferably is comprised double-stranded DNA.

An NOVX nucleic acid can encode a mature NOVX polypeptide. As used herein, a “mature” form of a polypeptide or protein disclosed in the present invention is the product of a naturally occurring polypeptide or precursor form or proprotein. The naturally occurring polypeptide, precursor or proprotein includes, by way of nonlimiting example, the full-length



gene product, encoded by the corresponding gene. Alternatively, it may be defined as the polypeptide, precursor or proprotein encoded by an ORF described herein. The product “mature” form arises, again by way of nonlimiting example, as a result of one or more naturally occurring processing steps as they may take place within the cell, or host cell, in which the gene product arises. Examples of such processing steps leading to a “mature” form of a polypeptide or protein include the cleavage of the N-terminal methionine residue encoded by the initiation codon of an ORF, or the proteolytic cleavage of a signal peptide or leader sequence. Thus a mature form arising from a precursor polypeptide or protein that has residues 1 to N, where residue 1 is the N-terminal methionine, would have residues 2 through N remaining after removal of the N-terminal methionine. Alternatively, a mature form arising from a precursor polypeptide or protein having residues 1 to N, in which an N-terminal signal sequence from residue 1 to residue M is cleaved, would have the residues from residue M+1 to residue N remaining. Further as used herein, a “mature” form of a polypeptide or protein may arise from a step of post-translational modification other than a proteolytic cleavage event. Such additional processes include, by way of non-limiting example, glycosylation, myristoylation or phosphorylation. In general, a mature polypeptide or protein may result from the operation of only one of these processes, or a combination of any of them.

The term “probes”, as utilized herein, refers to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), 100 nt, or as many as approximately, *e.g.*, 6,000 nt, depending upon the specific use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are generally obtained from a natural or recombinant source, are highly specific, and much slower to hybridize than shorter-length oligomer probes. Probes may be single- or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

The term “isolated” nucleic acid molecule, as utilized herein, is one, which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an “isolated” nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5'- and 3'-termini of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated NOVX nucleic acid molecules can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell/tissue from which the nucleic acid is derived (*e.g.*, brain, heart, liver, spleen, etc.). Moreover, an “isolated” nucleic acid molecule, such as a cDNA molecule, can

be substantially free of other cellular material or culture medium when produced by recombinant techniques, or of chemical precursors or other chemicals when chemically synthesized.

5 A nucleic acid molecule of the invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31, or a complement of this aforementioned nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31 as a hybridization probe, NOVX molecules can be isolated using standard
10 hybridization and cloning techniques (*e.g.*, as described in Sambrook, *et al.*, (eds.), MOLECULAR CLONING: A LABORATORY MANUAL 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively,
15 genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to NOVX nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

20 As used herein, the term "oligonucleotide" refers to a series of linked nucleotide residues, which oligonucleotide has a sufficient number of nucleotide bases to be used in a PCR reaction. A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue.
25 Oligonucleotides comprise portions of a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment of the invention, an oligonucleotide comprising a nucleic acid molecule less than 100 nt in length would further comprise at least 6 contiguous nucleotides SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31, or a complement thereof. Oligonucleotides may be
30 chemically synthesized and may also be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31, or a portion of this nucleotide sequence (*e.g.*, a fragment that can be used as a probe or primer or a fragment encoding a

biologically-active portion of an NOVX polypeptide). A nucleic acid molecule that is complementary to the nucleotide sequence shown SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, or 31 is one that is sufficiently complementary to the nucleotide sequence shown SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, or 31 that it
5 can hydrogen bond with little or no mismatches to the nucleotide sequence shown SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31, thereby forming a stable duplex.

As used herein, the term “complementary” refers to Watson-Crick or Hoogsteen base pairing between nucleotides units of a nucleic acid molecule, and the term “binding” means
10 the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, van der Waals, hydrophobic interactions, and the like. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the
15 effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

Fragments provided herein are defined as sequences of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of
20 amino acids, respectively, and are at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice. Derivatives are nucleic acid sequences or amino acid sequences formed from the native compounds either directly or by modification or partial substitution. Analogs are nucleic acid sequences or amino acid sequences that have a structure similar to, but not
25 identical to, the native compound but differs from it in respect to certain components or side chains. Analogs may be synthetic or from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type. Homologs are nucleic acid sequences or amino acid sequences of a particular gene that are derived from different species.

Derivatives and analogs may be full length or other than full length, if the derivative or
30 analog contains a modified nucleic acid or amino acid, as described below. Derivatives or analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the invention, in various embodiments, by at least about 70%, 80%, or 95% identity (with a preferred identity of 80-95%) over a nucleic acid or amino acid sequence of

identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the aforementioned proteins under stringent, moderately stringent, or low stringent conditions. *See e.g.* Ausubel, *et al.*, CURRENT
5 PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below.

A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences encode those sequences coding for isoforms of NOVX polypeptides. Isoforms can be expressed in different
10 tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the invention, homologous nucleotide sequences include nucleotide sequences encoding for an NOVX polypeptide of species other than humans, including, but not limited to: vertebrates, and thus can include, *e.g.*, frog, mouse, rat, rabbit, dog, cat, cow, horse, and other organisms. Homologous nucleotide
15 sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the exact nucleotide sequence encoding human NOVX protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17,
20 19, 21, 23, 25, 27, 29, and 31, as well as a polypeptide possessing NOVX biological activity. Various biological activities of the NOVX proteins are described below.

An NOVX polypeptide is encoded by the open reading frame ("ORF") of an NOVX nucleic acid. An ORF corresponds to a nucleotide sequence that could potentially be translated into a polypeptide. A stretch of nucleic acids comprising an ORF is uninterrupted by a stop
25 codon. An ORF that represents the coding sequence for a full protein begins with an ATG "start" codon and terminates with one of the three "stop" codons, namely, TAA, TAG, or TGA. For the purposes of this invention, an ORF may be any part of a coding sequence, with or without a start codon, a stop codon, or both. For an ORF to be considered as a good candidate for coding for a *bona fide* cellular protein, a minimum size requirement is often set,
30 *e.g.*, a stretch of DNA that would encode a protein of 50 amino acids or more.

The nucleotide sequences determined from the cloning of the human NOVX genes allows for the generation of probes and primers designed for use in identifying and/or cloning NOVX homologues in other cell types, *e.g.* from other tissues, as well as NOVX homologues from other vertebrates. The probe/primer typically comprises substantially purified

oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 consecutive sense strand nucleotide sequence SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, or 31; or an anti-sense strand nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, or 31; or of a naturally occurring mutant of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31.

Probes based on the human NOVX nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the probe further comprises a label group attached thereto, *e.g.* the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissues which mis-express an NOVX protein, such as by measuring a level of an NOVX-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting NOVX mRNA levels or determining whether a genomic NOVX gene has been mutated or deleted.

“A polypeptide having a biologically-active portion of an NOVX polypeptide” refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically-active portion of NOVX" can be prepared by isolating a portion SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, or 31, that encodes a polypeptide having an NOVX biological activity (the biological activities of the NOVX proteins are described below), expressing the encoded portion of NOVX protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of NOVX.

NOVX Nucleic Acid and Polypeptide Variants

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences shown in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31 due to degeneracy of the genetic code and thus encode the same NOVX proteins as that encoded by the nucleotide sequences shown in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32.

In addition to the human NOVX nucleotide sequences shown in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31, it will be appreciated by those skilled in the

art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the NOVX polypeptides may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the NOVX genes may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to
5 nucleic acid molecules comprising an open reading frame (ORF) encoding an NOVX protein, preferably a vertebrate NOVX protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the NOVX genes. Any and all such nucleotide variations and resulting amino acid polymorphisms in the NOVX polypeptides, which are the result of natural allelic variation and that do not alter the functional activity of the NOVX
10 polypeptides, are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding NOVX proteins from other species, and thus that have a nucleotide sequence that differs from the human SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31 are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of
15 the NOVX cDNAs of the invention can be isolated based on their homology to the human NOVX nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the
20 invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500, 750, 1000, 1500, or 2000 or more nucleotides in length. In yet another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the
25 coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other.

Homologs (*i.e.*, nucleic acids encoding NOVX proteins derived from species other than human) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or
30 high stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in

different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5 °C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess, at T_m, 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes, primers or oligonucleotides (*e.g.*, 10 nt to 50 nt) and at least about 60°C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions are hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65°C, followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequences SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31, corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31, or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55°C, followed by one or more washes in 1X SSC, 0.1% SDS at 37°C. Other conditions of moderate stringency that may be used are well-known within the art. *See, e.g.*, Ausubel, *et al.* (eds.), 1993, CURRENT PROTOCOLS IN

MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990; GENE TRANSFER AND
EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule
comprising the nucleotide sequences SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25,
27, 29, and 31, or fragments, analogs or derivatives thereof, under conditions of low
stringency, is provided. A non-limiting example of low stringency hybridization conditions
are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA,
0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10%
(wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl
(pH 7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other conditions of low stringency that may
be used are well known in the art (*e.g.*, as employed for cross-species hybridizations). *See,*
e.g., Ausubel, *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley
& Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY
MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981. *Proc Natl Acad Sci USA* 78:
6789-6792.

Conservative Mutations

In addition to naturally-occurring allelic variants of NOVX sequences that may exist in
the population, the skilled artisan will further appreciate that changes can be introduced by
mutation into the nucleotide sequences SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23,
25, 27, 29, and 31, thereby leading to changes in the amino acid sequences of the encoded
NOVX proteins, without altering the functional ability of said NOVX proteins. For example,
nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid
residues can be made in the sequence SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24,
26, 28, 30, or 32. A "non-essential" amino acid residue is a residue that can be altered from
the wild-type sequences of the NOVX proteins without altering their biological activity,
whereas an "essential" amino acid residue is required for such biological activity. For
example, amino acid residues that are conserved among the NOVX proteins of the invention
are predicted to be particularly non-amenable to alteration. Amino acids for which
conservative substitutions can be made are well-known within the art.

Another aspect of the invention pertains to nucleic acid molecules encoding NOVX
proteins that contain changes in amino acid residues that are not essential for activity. Such
NOVX proteins differ in amino acid sequence from SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17,
19, 21, 23, 25, 27, 29, and 31 yet retain biological activity. In one embodiment, the isolated

nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 45% homologous to the amino acid sequences SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, and 32.

Preferably, the protein encoded by the nucleic acid molecule is at least about 60% homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, and 32; more preferably at least about 70% homologous SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32; still more preferably at least about 80% homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32; even more preferably at least about 90% homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32; and most preferably at least about 95% homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32.

An isolated nucleic acid molecule encoding an NOVX protein homologous to the protein of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32 can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

Mutations can be introduced into SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted, non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined within the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted non-essential amino acid residue in the NOVX protein is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of an NOVX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for NOVX biological activity to identify mutants that retain activity. Following mutagenesis SEQ ID NOS:1, 3, 5, 7, 9, 11, 13,

15, 17, 19, 21, 23, 25, 27, 29, and 31, the encoded protein can be expressed by any
 recombinant technology known in the art and the activity of the protein can be determined.

The relatedness of amino acid families may also be determined based on side chain
 interactions. Substituted amino acids may be fully conserved "strong" residues or fully
 5 conserved "weak" residues. The "strong" group of conserved amino acid residues may be any
 one of the following groups: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW,
 wherein the single letter amino acid codes are grouped by those amino acids that may be
 substituted for each other. Likewise, the "weak" group of conserved residues may be any one
 of the following: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK,
 10 VLIM, HFY, wherein the letters within each group represent the single letter amino acid code.

In one embodiment, a mutant NOVX protein can be assayed for (i) the ability to form
 protein:protein interactions with other NOVX proteins, other cell-surface proteins, or
 biologically-active portions thereof, (ii) complex formation between a mutant NOVX protein
 and an NOVX ligand; or (iii) the ability of a mutant NOVX protein to bind to an intracellular
 15 target protein or biologically-active portion thereof; (*e.g.* avidin proteins).

In yet another embodiment, a mutant NOVX protein can be assayed for the ability to
 regulate a specific biological function (*e.g.*, regulation of insulin release).

Antisense Nucleic Acids

Another aspect of the invention pertains to isolated antisense nucleic acid molecules
 20 that are hybridizable to or complementary to the nucleic acid molecule comprising the
 nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and
 31, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a
 nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein (*e.g.*,
 complementary to the coding strand of a double-stranded cDNA molecule or complementary
 25 to an mRNA sequence). In specific aspects, antisense nucleic acid molecules are provided that
 comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides
 or an entire NOVX coding strand, or to only a portion thereof. Nucleic acid molecules
 encoding fragments, homologs, derivatives and analogs of an NOVX protein of SEQ ID
 NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32, or antisense nucleic acids
 30 complementary to an NOVX nucleic acid sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15,
 17, 19, 21, 23, 25, 27, 29, and 31, are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding
 region" of the coding strand of a nucleotide sequence encoding an NOVX protein. The term

"coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding the NOVX protein. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding the NOVX protein disclosed herein, antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of NOVX mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of NOVX mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of NOVX mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally-occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids (*e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used).

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the

inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding an NOVX protein to thereby inhibit expression of the protein (*e.g.*, by inhibiting transcription and/or translation). The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface (*e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens). The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient nucleic acid molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α -anomeric nucleic acid molecule. An α -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β -units, the strands run parallel to each other. *See, e.g., Gaultier, et al., 1987. Nucl. Acids Res. 15: 6625-6641.* The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (*See, e.g., Inoue, et al. 1987. Nucl. Acids Res. 15: 6131-6148*) or a chimeric RNA-DNA analogue (*See, e.g., Inoue, et al., 1987. FEBS Lett. 215: 327-330.*

Ribozymes and PNA Moieties

Nucleic acid modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

In one embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of

cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes as described in Haselhoff and Gerlach 1988. *Nature* 334: 585-591) can be used to catalytically cleave NOVX mRNA transcripts to thereby inhibit translation of NOVX mRNA. A ribozyme having
5 specificity for an NOVX-encoding nucleic acid can be designed based upon the nucleotide sequence of an NOVX cDNA disclosed herein (*i.e.*, SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in an NOVX-encoding mRNA. *See, e.g.*, U.S. Patent
10 4,987,071 to Cech, *et al.* and U.S. Patent 5,116,742 to Cech, *et al.* NOVX mRNA can also be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. *See, e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, NOVX gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the NOVX nucleic acid (*e.g.*, the NOVX
15 promoter and/or enhancers) to form triple helical structures that prevent transcription of the NOVX gene in target cells. *See, e.g.*, Helene, 1991. *Anticancer Drug Des.* 6: 569-84; Helene, *et al.* 1992. *Ann. N.Y. Acad. Sci.* 660: 27-36; Maher, 1992. *Bioassays* 14: 807-15.

In various embodiments, the NOVX nucleic acids can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility
20 of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids. *See, e.g.*, Hyrup, *et al.*, 1996. *Bioorg Med Chem* 4: 5-23. As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics (*e.g.*, DNA mimics) in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral
25 backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup, *et al.*, 1996. *supra*; Perry-O'Keefe, *et al.*, 1996. *Proc. Natl. Acad. Sci. USA* 93: 14670-14675.

PNAs of NOVX can be used in therapeutic and diagnostic applications. For example,
30 PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of NOVX can also be used, for example, in the analysis of single base pair mutations in a gene (*e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S₁ nucleases (*See, Hyrup, et al.*, 1996.*supra*); or as probes or primers

for DNA sequence and hybridization (*See*, Hyrup, *et al.*, 1996, *supra*; Perry-O'Keefe, *et al.*, 1996. *supra*).

In another embodiment, PNAs of NOVX can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of NOVX can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes (*e.g.*, RNase H and DNA polymerases) to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (*see*, Hyrup, *et al.*, 1996. *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup, *et al.*, 1996. *supra* and Finn, *et al.*, 1996. *Nucl Acids Res* 24: 3357-3363. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA. *See, e.g.*, Mag, *et al.*, 1989. *Nucl Acid Res* 17: 5973-5988. PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment. *See, e.g.*, Finn, *et al.*, 1996. *supra*. Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. *See, e.g.*, Petersen, *et al.*, 1975. *Bioorg. Med. Chem. Lett.* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (*see, e.g.*, Letsinger, *et al.*, 1989. *Proc. Natl. Acad. Sci. U.S.A.* 86: 6553-6556; Lemaitre, *et al.*, 1987. *Proc. Natl. Acad. Sci.* 84: 648-652; PCT Publication No. WO88/09810) or the blood-brain barrier (*see, e.g.*, PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (*see, e.g.*, Krol, *et al.*, 1988. *BioTechniques* 6:958-976) or intercalating agents (*see, e.g.*, Zon, 1988. *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, and the like.

NOVX Polypeptides

A polypeptide according to the invention includes a polypeptide including the amino acid sequence of NOVX polypeptides whose sequences are provided in SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residues shown in SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32 while still encoding a protein that maintains its NOVX activities and physiological functions, or a functional fragment thereof.

In general, an NOVX variant that preserves NOVX-like function includes any variant in which residues at a particular position in the sequence have been substituted by other amino acids, and further include the possibility of inserting an additional residue or residues between two residues of the parent protein as well as the possibility of deleting one or more residues from the parent sequence. Any amino acid substitution, insertion, or deletion is encompassed by the invention. In favorable circumstances, the substitution is a conservative substitution as defined above.

One aspect of the invention pertains to isolated NOVX proteins, and biologically-active portions thereof, or derivatives, fragments, analogs or homologs thereof. Also provided are polypeptide fragments suitable for use as immunogens to raise anti-NOVX antibodies. In one embodiment, native NOVX proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, NOVX proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, an NOVX protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" polypeptide or protein or biologically-active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the NOVX protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of NOVX proteins in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly-produced. In one embodiment, the language "substantially free of cellular material" includes preparations of NOVX proteins having less than about 30% (by dry weight) of non-NOVX proteins (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-NOVX proteins, still more preferably less than about 10% of non-NOVX proteins, and most preferably less than about 5% of non-NOVX proteins. When the NOVX protein or

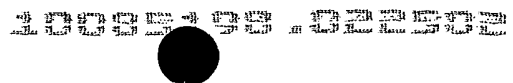
biologically-active portion thereof is recombinantly-produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the NOVX protein preparation.

5 The language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX proteins in which the protein is separated from chemical precursors or other chemicals that are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX proteins having less than about 30% (by dry weight) of chemical precursors or
10 non-NOVX chemicals, more preferably less than about 20% chemical precursors or non-NOVX chemicals, still more preferably less than about 10% chemical precursors or non-NOVX chemicals, and most preferably less than about 5% chemical precursors or non-NOVX chemicals.

 Biologically-active portions of NOVX proteins include peptides comprising amino
15 acid sequences sufficiently homologous to or derived from the amino acid sequences of the NOVX proteins (*e.g.*, the amino acid sequence shown in SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32) that include fewer amino acids than the full-length NOVX proteins, and exhibit at least one activity of an NOVX protein. Typically, biologically-active portions comprise a domain or motif with at least one activity of the NOVX protein. A
20 biologically-active portion of an NOVX protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acid residues in length.

 Moreover, other biologically-active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native NOVX protein.

25 In an embodiment, the NOVX protein has an amino acid sequence shown SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32. In other embodiments, the NOVX protein is substantially homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32, and retains the functional activity of the protein of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32, yet differs in amino acid sequence due to
30 natural allelic variation or mutagenesis, as described in detail, below. Accordingly, in another embodiment, the NOVX protein is a protein that comprises an amino acid sequence at least about 45% homologous to the amino acid sequence SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32, and retains the functional activity of the NOVX proteins of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32.



Determining Homology Between Two or More Sequences

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. *See*, Needleman and Wunsch, 1970. *J Mol Biol* 48: 443-453. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence shown in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31.

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region.

Chimeric and Fusion Proteins

The invention also provides NOVX chimeric or fusion proteins. As used herein, an NOVX "chimeric protein" or "fusion protein" comprises an NOVX polypeptide operatively-
 5 linked to a non-NOVX polypeptide. An "NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to an NOVX protein SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, or 32, whereas a "non-NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the NOVX protein, *e.g.*, a protein that is different from the NOVX protein and
 10 that is derived from the same or a different organism. Within an NOVX fusion protein the NOVX polypeptide can correspond to all or a portion of an NOVX protein. In one embodiment, an NOVX fusion protein comprises at least one biologically-active portion of an NOVX protein. In another embodiment, an NOVX fusion protein comprises at least two biologically-active portions of an NOVX protein. In yet another embodiment, an NOVX
 15 fusion protein comprises at least three biologically-active portions of an NOVX protein. Within the fusion protein, the term "operatively-linked" is intended to indicate that the NOVX polypeptide and the non-NOVX polypeptide are fused in-frame with one another. The non-NOVX polypeptide can be fused to the N-terminus or C-terminus of the NOVX polypeptide.

20 In one embodiment, the fusion protein is a GST-NOVX fusion protein in which the NOVX sequences are fused to the C-terminus of the GST (glutathione S-transferase) sequences. Such fusion proteins can facilitate the purification of recombinant NOVX polypeptides.

In another embodiment, the fusion protein is an NOVX protein containing a
 25 heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of NOVX can be increased through use of a heterologous signal sequence.

In yet another embodiment, the fusion protein is an NOVX-immunoglobulin fusion protein in which the NOVX sequences are fused to sequences derived from a member of the
 30 immunoglobulin protein family. The NOVX-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between an NOVX ligand and an NOVX protein on the surface of a cell, to thereby suppress NOVX-mediated signal transduction *in vivo*. The NOVX-immunoglobulin fusion proteins can be used to affect the bioavailability of an NOVX cognate ligand.

Inhibition of the NOVX ligand/NOVX interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, as well as modulating (*e.g.* promoting or inhibiting) cell survival. Moreover, the NOVX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-NOVX antibodies in a subject, to purify NOVX ligands, and in screening assays to identify molecules that inhibit the interaction of NOVX with an NOVX ligand.

An NOVX chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (*see, e.g.*, Ausubel, *et al.* (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). An NOVX-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the NOVX protein.

NOVX Agonists and Antagonists

The invention also pertains to variants of the NOVX proteins that function as either NOVX agonists (*i.e.*, mimetics) or as NOVX antagonists. Variants of the NOVX protein can be generated by mutagenesis (*e.g.*, discrete point mutation or truncation of the NOVX protein). An agonist of the NOVX protein can retain substantially the same, or a subset of, the biological activities of the naturally occurring form of the NOVX protein. An antagonist of the NOVX protein can inhibit one or more of the activities of the naturally occurring form of the NOVX protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the NOVX protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities

of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the NOVX proteins.

5 Variants of the NOVX proteins that function as either NOVX agonists (*i.e.*, mimetics) or as NOVX antagonists can be identified by screening combinatorial libraries of mutants (e.g., truncation mutants) of the NOVX proteins for NOVX protein agonist or antagonist activity. In one embodiment, a variegated library of NOVX variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of NOVX variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a
10 degenerate set of potential NOVX sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of NOVX sequences therein. There are a variety of methods which can be used to produce libraries of potential NOVX variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate
15 set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential NOVX sequences. Methods for synthesizing degenerate oligonucleotides are well-known within the art. *See, e.g.*, Narang, 1983. *Tetrahedron* 39: 3; Itakura, *et al.*, 1984. *Annu. Rev. Biochem.* 53: 323; Itakura, *et al.*, 1984. *Science* 198: 1056; Ike, *et al.*, 1983. *Nucl. Acids Res.* 11: 477.

Polypeptide Libraries

In addition, libraries of fragments of the NOVX protein coding sequences can be used to generate a variegated population of NOVX fragments for screening and subsequent
25 selection of variants of an NOVX protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of an NOVX coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double-stranded DNA that can include sense/antisense pairs from different nicked products, removing single
30 stranded portions from reformed duplexes by treatment with S₁ nuclease, and ligating the resulting fragment library into an expression vector. By this method, expression libraries can be derived which encodes N-terminal and internal fragments of various sizes of the NOVX proteins.

Various techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of NOVX proteins. The most
5 widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble
10 mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify NOVX variants. See, e.g., Arkin and Yourvan, 1992. *Proc. Natl. Acad. Sci. USA* 89: 7811-7815; Delgrave, et al., 1993. *Protein Engineering* 6:327-331.

Anti-NOVX Antibodies

Also included in the invention are antibodies to NOVX proteins, or fragments of NOVX proteins. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, i.e., molecules that contain an antigen binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab},
20 F_{ab'} and F_{(ab')₂} fragments, and an F_{ab} expression library. In general, an antibody molecule obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG₁, IgG₂, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a
25 reference to all such classes, subclasses and types of human antibody species.

An isolated NOVX-related protein of the invention may be intended to serve as an antigen, or a portion or fragment thereof, and additionally can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for polyclonal and monoclonal antibody preparation. The full-length protein can be used or,
30 alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length